

Kyoto for commuters

Offset schemes are a small but potentially useful addition to the carbon balance sheet.

Climate change, as we are now experiencing it, is predominantly the result of the lifestyles to which people living in the industrialized world have grown accustomed. One increasingly popular response to it is carbon offsetting, a practice that allows people to compensate for the impact of their activities by supporting climate-friendly projects in distant corners of the world. It is a promising approach — even if some of its early manifestations are being dangerously oversold (see page 976).

Some of the more popular carbon offsetting schemes allow consumers to buy credits on an emerging, but unregulated, voluntary emissions market, to compensate for carbon-emitting behaviour such as air travel, commuting or home heating. Outfits such as the CarbonNeutral Company in London or Atmosfair in Bonn will sell such credits for a few euros, or tens of euros. The money will ultimately be used, they say, to reduce carbon emissions in regions of the world where it is inexpensive to do so: for example, households in India or Peru might switch to using their own locally produced biofuels for cooking.

Industrial companies and organizers of large events, such as the 2006 World Cup, are also able to buy credits for larger-scale energy projects to offset their emissions in the same way.

The principle of carbon offsetting makes sense because it makes no difference to the atmosphere where carbon emissions are reduced, as long as they happen. It has rapidly become quite fashionable in some countries, notably Britain and Germany, where an improbable alliance of green groups, travel agencies and corporations have rallied behind it. Global concern about climate change, it would appear, is creating a new breed of responsible consumer, whose participation may serve to reduce global pollution and poverty.

So much goodwill at once is arousing some suspicion. But the idea should not be dismissed out of hand. At its best, voluntary carbon offsetting is an elegant way of mixing capitalism and sentiment in the fight against global warming.

The voluntary market has many problems, however, and not every offset project will deliver the emissions cuts promised. More transparency and better regulation would make offsetting schemes more attractive to those who remain sceptical of their value. The United

Nations Climate Secretariat, which oversees mandatory emissions trading under the Kyoto Protocol, and national bodies such as the UK Financial Services Authority, need to work together on the accreditation of schemes that participate in the emerging, voluntary market, and the regulation of the way they are sold to consumers. Yet it is important that the accreditation is not so heavy-handed that it kills off the small, inexpensive projects that need to materialize if carbon offsetting is to work.

Until such accreditation is in place, the idea should not be oversold. The cuts that can be achieved in this way in poor nations are somewhat limited in scope, and encouraging them should not be seen as a substitute for the political action needed to reduce emissions in industrialized countries. Consumption, mobility, global tourism and many other aspects of modern life are going to have to change if global warming is to be confronted effectively.

Unwillingness to acknowledge that existing lifestyle choices may be incompatible with the need to confront the problem is a theme that comes up repeatedly in climate debates. After all, it is not just George W. Bush, Tony Blair and Angela Merkel who are failing to adequately address the issue. Consumers of goods and services that depend on the burning of fossil fuels also have a role to play.

There is obviously a danger that the procurement of credits on such a scheme may induce self-satisfaction, complacency and perhaps even extravagance on the part of a few individuals. However, this effect is small compared with the general raising of public consciousness on the issue of global warming that is encouraged by the high profile of the offsetting plans.

Carbon offsetting is becoming popular as individuals and businesses begin to recognize this. Whether they are acting for ethical reasons, or simply in response to the preferences of investors, customers and employees, the final effect remains beneficial — provided that the low-carbon projects that get a boost as a result are properly devised, implemented and monitored. ■

"At its best, voluntary carbon offsetting is an elegant way of mixing capitalism and sentiment in the fight against global warming."

Peer review and fraud

Two assessments of the refereeing process highlight challenges for journals.

It would be misguided indeed for *Nature* to have any competitor's sense of Schadenfreude over *Science*'s experiences with two papers on human embryonic stem cells by Woo Suk Hwang and his colleagues. It is possible that we at *Nature* would have published

the papers had they been submitted to us instead.

The external review committee, which included *Nature*'s US executive editor Linda Miller, earlier this month endorsed the general level of care exercised by *Science*. It also stated that, cumulatively, concerns raised during the refereeing processes should have been pursued more sceptically. Hindsight is always 20/20, however, and it is not at all clear that the questions were sufficiently pressing that any journal would have pursued them, given the degree of trust that is required between authors and editors.

It is precisely this degree of trust that the review panel goes on to



challenge. Given the fact that cases of fraud demonstrably make it through refereeing processes, and given the importance of public trust in science, it proposes that journals apply additional scrutiny and risk assessment to papers that are likely to have a significant public impact, such as those with direct implications for policy, public health or climate change. The additional scrutiny recommended by the panel includes greater attention to raw data and a clarification of the contributions of each co-author.

We at *Nature*, like the editors at *Science*, accept that this challenge has to be addressed, and we have accepted their invitation to deal with such matters collaboratively. The key is to raise editors' and referees' practical awareness of the risk of deception. A conscious risk assessment, in which the likelihood of deception is explicitly analysed, is much easier said than done. An elementary checklist of risk factors can readily be introduced. But it would, for example, be quite inappropriate to single out papers from Seoul National University, or on human embryonic stem cells, and apply higher thresholds of proof to them.

On the other hand, we do already seek to ensure that major claims are backed by rigorous data and argument. Nevertheless, the sad fact is that high-profile and problematic papers have occasionally slipped through the net, and we accept that this underlines the need for enhanced attention by editors.

Nature and the Nature research journals already encourage the specification of authors' contributions to papers, and the uptake of this by authors has increased greatly in the past year — a fact that is welcomed by some funding agencies. We now intend to conduct a survey to help us decide whether to make this practice

compulsory, and we would welcome readers' feedback.

It is an interesting question whether a more open peer-review process might have led to the detection of Hwang's fraud. At present, however, the level of interest in open peer review is too small to hope for such an outcome. That, at least, is the implication of *Nature's* trial of open peer review, the results of which can be found in an online debate on the subject (www.nature.com/nature/peerreview/debate/nature05535.html).

In the trial, the papers selected for traditional peer review were, in a parallel option offered to authors, hosted for public comment. In the event, 5% of authors took up this option. Although most authors found at least some value in the comments they received, they were small in number, and editors did not think they contributed significantly to their decisions.

This was not a controlled experiment, so in no sense does it disprove the hypothesis that open peer review could one day become accepted practice. But this experience, along with informal discussions with researchers, suggests that most of them are too busy, and lack sufficient career incentive, to venture onto a venue such as *Nature's* website and post public, critical assessments of their peers' work.

Another form of peer review emerges after publication, when work is replicated — or not. If this kind of discussion is to make it into the open, rather than be confined to gossip at conferences, it requires a forum where peers are able to comment on individual papers, with minimal editorial intervention. Would commenting on *Nature* papers be more widely adopted by researchers after they have been formally published than before? We intend to introduce this function next year, and find out. ■

Days of Futures past

It's a time of change for *Nature's* venture into speculative fiction.

This week features the last in the series of 'Futures', *Nature's* occasional excursion into speculative fiction (as science fiction, or SF, is sometimes known). Since 1999, *Nature* has published no fewer than 156 such journeys into things to come, ranging from the serious to the whimsical, all of which (hopefully) provided some entertainment.

And that's the key — SF is meant to amuse in the present. The most memorable SF works do this by projecting our present concerns onto the greater canvas of uncertainty that is the preserve of that part of history that hasn't happened yet. Or, to put it more succinctly, "it is difficult to make predictions, especially about the future", a sentiment usually attributed to baseball star Yogi Berra. That it might sometimes be found attached to Sam Goldwyn, Woody Allen or even Niels Bohr demonstrates its fundamental soundness — and also, perhaps, that the past is as unreliable as the future.

Addressing this very issue, one visionary bucked the trend in the grandest fashion, writing in 1902 that "the time is drawing near when it will be possible to suggest a systematic exploration of the future". This is a bold statement indeed, but one that makes more sense if it

is recognized that science is a necessarily predictive endeavour. As if to prove his point, the writer went on to outline a startling version of statistical mechanics in which the behaviour of large numbers of humans could be predicted, presaging Isaac Asimov's future science of psychohistory (in his 'Foundation' novels) by half a century.

Later still, the writer notes that "the question what is to come after man is the most persistently fascinating and the most insoluble question in the whole world", touching, even if coincidentally, on the theme of post-humanity that animates much current SF, itself a reaction to cloning and genetic manipulation.

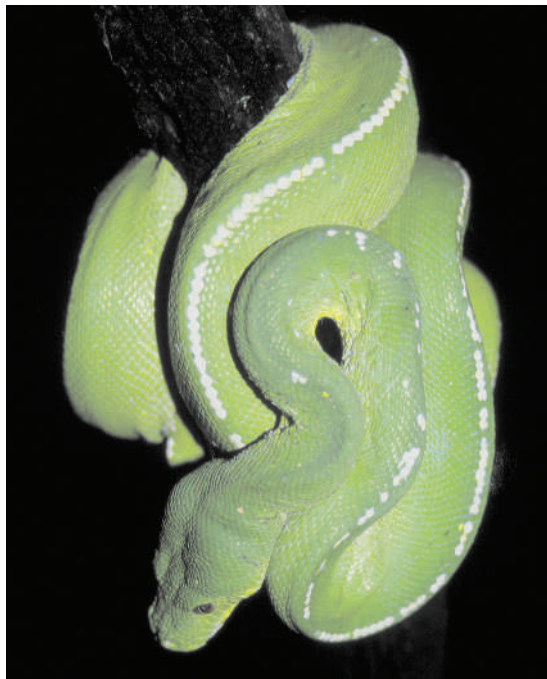
Who was this visionary? It was none other than H. G. Wells, writing in these very pages (*Nature* 65, 326–331; 1902). From this we can derive some comfort that 'Futures' in *Nature* has indeed a venerable past.

And even now, we haven't quite come to the end of eternity. An anthology of the best of Futures is planned, and the column itself will continue, for a while, in our sister publication *Nature Physics*. So for a little longer, at least, we can share the emotions of Wells' Time Traveller (in his 1895 novel *The Time Machine*), who "merged at last into a kind of hysterical exhilaration... with a kind of madness growing upon me, I flung myself into futurity". ■

"An anthology of the best of Futures is planned, and the column itself will continue, for a while, in *Nature Physics*."

RESEARCH HIGHLIGHTS

D. WILSON

**Hunting attire***Biol. Lett.* doi:10.1098/rsbl.2006.0574 (2006)

A study of the tropical python (*Morelia viridis*, pictured) explains why these snakes are born yellow or brick-red in colour, then become green as they grow up.

Researchers led by Robert Heinsohn of the Australian National University in Canberra analysed how the different hues appear to the visual systems of the snakes' main predators, birds. They found that yellow and red are the least conspicuous colours on the ground at the forest's edge, where juvenile snakes hunt. But once the snakes are big enough to prey on rodents and small birds, they head into the forest canopy in search of a meal. The snake's colour switch is linked to this shift in foraging ground: in the canopy, green is better camouflage.

MICROBIOLOGY**A break from the norm***Nature Immunol.* doi:10.1038/ni1423 (2006)

When the bacterium *Shigella flexneri* invades the colon and triggers dysentery, it is thought to modify the normal inflammatory response of cells lining the intestine. Laurence Arbibe at the Pasteur Institute in Paris, France, and her colleagues show that it does this by altering part of the cells' epigenetic information — that found outside the genetic code.

The histone H3 protein (which helps package up DNA) partly controls the ability of the protein NF- κ B to access and switch on a set of inflammatory genes that fend off pathogens. A *Shigella* protein called OspF interferes with this process. It prevents the normal addition of a phosphate group to histone H3, and by doing so blocks activation of these genes by NF- κ B and represses inflammation.

CLIMATE SCIENCE**Under a cloud***Proc. Natl Acad. Sci. USA* **103**, 19668–19672 (2006)

Cutting back on fossil fuel and biomass burning could boost India's crops, says the team behind an agro-economic model of the country's rice yields.

Fuel and biomass burning in India produces brown clouds that reflect the Sun's rays and partially shield the country from the effects of global warming. This has led to suggestions that reducing the burning might cause an acceleration in current temperature

increases that would damage rice yields.

But a computer simulation by Veerabhadran Ramanathan of the University of California, San Diego, and his colleagues shows that cutting back on fuel and biomass burning over the period of 1985–1998 would have led to a 10% increase in yields. The rise is due to the increase in rainfall predicted to occur when biomass pollution is reduced.

CANCER BIOLOGY**Genetic liability***PLoS Med.* **3**, e516 (2006)

A family that is prone to pancreatic cancer carries a mutation in a gene that may also play a role in non-inherited pancreatic tumours, report Teresa Brentnall of the University of Washington, Seattle, and her colleagues.

The researchers identified a gene, called *palladin*, that was mutated in 18 family members with pancreatic cancer or precancerous lesions, but not in family members free of the disease. They also showed that *palladin* is overexpressed in non-inherited pancreatic cancer and pre-cancer.

The *palladin* gene plays a role in building the cellular skeleton, and the team showed that cells with the mutated

form of the gene in human cell culture were 50% more mobile than normal cells. This motility might aid the tumour's invasion of other tissues.

BIOCHEMISTRY**Cycling backwards***J. Am. Chem. Soc.* doi:10.1021/ja066103k (2006)

Minerals could have stood in for enzymes to produce complex organic compounds on the early Earth, say Xiang Zhang and Scot Martin of Harvard University in Cambridge, Massachusetts.

The researchers show that zinc sulphide — present naturally as the mineral sphalerite — can act as a photocatalyst for three of the five key reductive reactions in the reverse Krebs cycle. Run forwards, the Krebs cycle forms a crucial component of our own metabolism. Aerobic cells use enzymes to drive a set of reactions that break down carbohydrates into carbon dioxide and generate energy.

The reverse Krebs cycle uses an energy source to build up such molecules from carbon dioxide. If all the reverse cycle's steps can be catalysed by minerals, it could have produced carbohydrates before enzymes even evolved.

ASTRONOMY**Catch a comet***Science* **314**, 1711–1739 (2006)

Detailed analyses of the comet dust brought back by NASA's Stardust spacecraft in January were published in a series of seven papers in *Science* last week. They reveal the spacecraft's target, comet 81P/Wild 2, to contain a vast

array of different particles — some from the heart of the Solar System, some from pre-solar times.

Among the cometary grains that had embedded themselves in the spacecraft's aerogel collector (particle track pictured, left), the researchers discovered one particle enriched in the isotope oxygen-17. This suggests that it came from a dying star that existed long before our Solar System. They also found minerals containing lots of calcium and aluminium, which must have formed in the hottest part of the collapsing gas cloud that gave rise to the Solar System. How these grains made the long journey



SCIENCE

from the centre of the Solar System to its icy edge where the comet is thought to have formed is still under discussion.

NEUROBIOLOGY

Brain, heal thyself

Cell 127, 1253–1264 (2006)

Certain regions of the brain can use their own stem cells to heal local damage, suggests a study in mice.

Yuh-Nung Jan of the University of California, San Francisco, and his colleagues looked at an area of the brain known as the subventricular zone (SVZ). They showed that a gene called *Numb* regulates how stem cells from the SVZ become neurons, and instructs these cells to maintain the walls of the lateral ventricles, the brain's central cavities.

With *Numb* knocked out, mice developed large holes in these walls. But, rather than worsening over time, the holes were repaired within 6 weeks. The team suggests that stem cells that escaped the knockout were able to shore up the walls. Such capacity for do-it-yourself repair might be harnessed to treat brain damage.

CLIMATE CHANGE

Predicting sea-level rise

Science doi:10.1126/science.1135456 (2006)

Future sea-level change is hard to predict, because the processes that drive it, such as glacial and icecap melting, are not easily modelled. Stefan Rahmstorf of the Potsdam Institute for Climate Impact Research in Germany now provides a semi-empirical approach to the problem.

Using simple mathematical assumptions and 120 years of historical data, he finds a relationship between the amount that temperature increases above a set baseline and the rate at which sea level rises. Applying this technique to projected temperature increases suggests that sea levels in 2100 will be between 0.5 and 1.4 metres higher than they were in 1990. Although this uncertainty may be greater than expected, the result indicates that very small sea-level rises look implausible.

PHYSICS

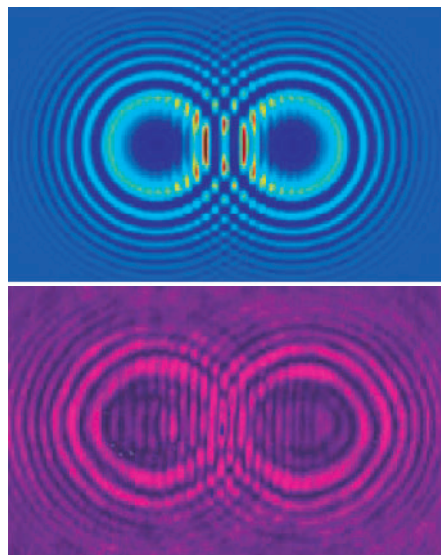
Ripple effect

Nature Phys. doi:10.1038/nphys486 (2006)

One route to good results is finding a convenient system to work with, as Jason Fleischer of Princeton University, New Jersey, and his colleagues demonstrate. They use an optical set-up to probe a kind of shock wave associated with superfluids.

The team created shock waves that mimic

those in superfluids by shining laser beams into a nonlinear crystal. The light behaves like a density wave in a liquid. Unlike shock waves in classical fluids, which have a well-defined shock front, those in superfluids — where flow is unhindered by viscosity — disperse into a series of ripples. Fleischer's group used their simple set-up to study how these ripples interact when two shock waves collide (pictured below; blue from



simulations, pink from experiments). This extends previous observations of Bose–Einstein condensates, a type of superfluid.

CHEMICAL SYNTHESIS

A big hand

J. Am. Chem. Soc. doi:10.1021/ja067840j (2006)

Making just one of two mirror-image (chiral) molecules is a great challenge for organic chemists at the best of times. Such molecules contain at least one 'stereogenic centre', typically a carbon atom around which four different substituents can be arranged in left- or right-handed orientations.

Scott Miller of Yale University in New Haven, Connecticut, and his colleagues tackle a harder problem still — finding a catalyst that can create a chiral molecule by transforming a part of the target molecule that sits half a nanometre from the stereogenic centre, while leaving an equivalent group untouched. An enzyme might be big enough to 'feel' the stereogenic centre from the remote reaction site, but Miller's group searched peptide libraries for smaller molecules that can do so too. Their best catalyst was a peptide containing just four amino acids. It's not clear how it works, but it gave a 95% excess of the desired chiral product.

JOURNAL CLUB

David R. Liu
Harvard University, Cambridge,
Massachusetts, USA

Thanks to the discovery of a new catalytic RNA, a chemical biologist can satisfy his student's curiosity.

An first-year undergraduate recently asked me a remarkable question: are all natural ribozymes, RNA molecules with catalytic activity, simply leftovers from the 'RNA world'? The RNA-world hypothesis supposes that RNA molecules were precursors to the first primitive forms of life, before the evolution of DNA and proteins.

Unanswered, my student's question preoccupied me until I encountered a recent paper by Jack Szostak of Massachusetts General Hospital, Boston, and his co-workers (K. Salehi-Ashtiani *et al.* *Science* 313, 1788–1792; 2006).

Prior to this work, only two of the known natural ribozymes were associated with mammals. The rarity of catalytic RNAs in more recently evolved, higher-order cells could reflect their attrition on the evolutionary battlefield during the rise of more highly functional protein enzymes.

This paper, however, supports a different conclusion. Szostak's group designed an ingenious system to isolate self-cleaving RNA molecules from RNA encoded in the human genome. Using this system, they discovered several new ribozymes.

One of these ribozymes, associated with a gene known as *CPEB3*, is highly conserved among placental mammals and marsupials, but is absent from non-mammalian vertebrates. This observation suggests that it arose relatively recently, around 200 million years ago.

We can therefore infer that some ribozymes have evolved in modern organisms, long after the era of the RNA world. The work elegantly demonstrates a new approach to the study of ancient molecules — and also reminds me that our youngest students can ask some of the best questions.

SPECIAL REPORT

Climate credits

Why change your lifestyle when you can pay a company to save your greenhouse-gas emissions for you? **Quirin Schiermeier** investigates whether carbon offsetting can really save the planet.

J. DECROW/AP

The 2006 World Cup in Germany will be remembered for France's Zinedine Zidane head-butting Italy's Marco Materazzi in the final. But in another sense, the largest sports event ever has passed without a trace: it has supposedly left no mark on the planet's atmosphere.

The international football association FIFA says it has met its 'Green Goal' to wipe out the World Cup's carbon footprint. It's estimated that the millions of fans who travelled to Germany to watch the matches generated around 100,000 tonnes of carbon dioxide. So local organizers collected €1.2 million (US\$1.6 million) from sponsors and used it to buy credits worth that amount on the voluntary emission market. The money will be invested in renewable-energy projects in developing countries.

Such voluntary carbon offsetting is becoming ever more fashionable: the 2012 Olympic Games in London will be labelled carbon neutral, and bands from Pink Floyd to Pearl Jam claim to rock carbon-free. British diplomats and government members jet around the world in a supposedly climate-friendly manner, as do many bankers and insurance brokers. Nicholas Stern, chief author of a recent report on the costs of climate change (see *Nature* 444, 6–7; 2006), is offsetting the CO₂ produced by his current promotional tour.

Offsetting is also becoming more popular among households, air travellers and car owners. There are currently around 40 retailers in Europe, Australia and North America that offer to save emissions on customers' behalf. And many travel companies allow customers to fund environmental projects worldwide.

It seems easy: for a small sum we can assuage our climate guilt — without changing our lifestyles. But does such offsetting really save the carbon emissions it promises? And can we really counter climate change by paying for carbon-saving projects elsewhere?

The projects on offer vary hugely. The first



Pearl Jam: aiming to be carbon neutral.

problem is simply calculating the amount of carbon that needs to be offset. For example, the British Carbon Neutral Company calculates a return flight from London to Bangkok, Thailand, at 2.1 tonnes of CO₂ per passenger, which it charges around €30 to offset. Swiss-based myclimate arrives at 3.6 tonnes and €86 for the same flight, and the German Atmosfair

reckons 6.9 tonnes and €139.

The figures differ because the Carbon Neutral Company calculates only the extra CO₂ emitted per passenger on a given route. Other providers multiply that impact by a factor of 2–3. This is because aircraft emissions, including nitric oxide, nitrogen dioxide and planes' vapour trails, have a more complex effect on clouds, ozone and climate than do those from earthbound polluters.

Hot air?

And if it's difficult to measure an individual's carbon footprint, it's almost impossible for big companies and events to decide which activities to include and where to draw the line between necessary and avoidable emissions, says Martin Cames, an emissions-trading expert with the Öko-Institut in Berlin. This means that extra emissions are often underestimated. For example, the Öko-Institut worked out the carbon footprint of the World Cup for FIFA, but simplified the calculation by counting only



the extra flights caused by the event.

Another issue is the choice of project. The cost of offsetting one tonne of CO₂ currently ranges from €3 to €30, and the projects on offer vary in type, scale and quality, from small fuel-switching projects to large-scale changes in land use and forest growth (see 'Setting off').

Buyers, private or corporate, need to make sure that the emission cuts that they pay for would not otherwise occur. If, for example, a measure is already required by law, the 'additionality' of a project, compared with business as usual, is not guaranteed. Sascha Lafeld, of the Frankfurt-based company Climate Change Consulting (3C), says the key is to check that the projects on offer require the external funding.

Setting off

Some of the ways consumers and companies are pegging back their emissions.

TerraPass

Offsetting agency that allows customers to offset emissions from cars, flights and homes. Money is ploughed into renewables and energy-

efficiency projects; customers can even exchange their old mobile phone for credit.

myclimate

By investing in projects from solar greenhouses in the Himalayas to methane power in South Africa, this agency is helping clients produce

carbon-neutral products, from cut flowers to fruit smoothies.

Ecoinsurance

Besides insuring British motorists, this company pledges to offset 20% of each customer's emissions by investing in renewable energy schemes.

**HERE'S TO 2006**

Find an end-of-year round-up and top-10 lists in our online special.

www.nature.com/news/specials/2006



according to a recent analysis by ICF, a London-based climate consultancy.

Russ George, founder and president of the offsetting company Planktos, based in Foster City, California, is one of those thinking big. He plans a 10,000-hectare 'climate park' in Hungary. Plant ecologists with the Hungarian Academy of Sciences have conducted a feasibility study of the environmental impact and permanence of the project, and planting is to begin in 2007. Planktos is also offering credits for its controversial ocean eco-restoration projects, which involve fertilizing plankton growth with iron.

Bottom line

George believes voluntary offsetting could solve the problem of climate change. "We're expecting to save the world and make a little money on the side," he says. "In that order."

Most experts agree that voluntary carbon offsetting could be part of a global strategy against climate change. But, they warn, most voluntary offsetting companies use small projects in developing countries, which together could account for only 3–5% of carbon emitted.

Worse, overselling offsetting might persuade consumers that no other action is needed.

"There's really nothing you can say against offsetting unavoidable emissions with truly permanent projects," says Olav Hohmeyer, an economist and renewable-energy expert at the University of Flensburg in Germany. "More importantly, however, consumers must question their behaviour."

"The symbolic value is certainly high," adds Ottmar Edenhofer, an economist at the Potsdam Institute for Climate Impact Research in Germany. "But it's nonsense to believe that a little bit of goodwill is all it takes to solve the problem."

Offsetting can deal with remaining emissions once avoidable energy consumption has been eliminated, he says. But most experts agree that the incentive to reduce consumption will come only when all sectors of business and industry are included in mandatory emissions-trading schemes.

If they were, industry would have to reduce emissions, and consumers would change their behaviour as the price of carbon rose. For example, Edenhofer estimates that including aviation in EU emissions-trading schemes would make flights 50–80% more expensive. Only such a move would force people to reconsider flying, he says: "Carbon dioxide must have a price — full stop."

Additional reporting by Michael Hopkin.

Up in the air: different companies give hugely varying estimates of the cost of offsetting a flight.

Also important is ensuring that offset projects are not counted more than once. This might happen if the same credit is sold to more than one buyer, or if voluntary reductions are counted against mandatory national targets.

And even if a project looks good on paper, such deals are risky. Most are located in developing countries, which don't have emission targets set by the Kyoto Protocol. But credits are usually sold upfront, with a promise that the money will be put into future projects, so there is a significant risk that the projects could fail to deliver reductions, or to materialize at all.

Bull market

The market's main weakness is a lack of standards and verification procedures. But guidelines are available from the UK Carbon Trust and the Clean Air–Cool Planet initiative in the United States. And the UK government's Environmental Audit Committee launched an inquiry last week into the issues surrounding voluntary offsetting. One key question is whether such schemes should require official accreditation.

The most widely recognized certification so far, called the Gold Standard, is owned and

supported by 42 non-governmental environmental organizations. It accepts only projects with proven additionality, and that have social and environmental benefits for local communities. Afforestation and reforestation projects are excluded, mainly because there is no guarantee that a forest will be permanent. When trees die, they release all the CO₂ they absorbed during their lives.

So, if we can calculate our emissions and invest in the right schemes, could offsetting save the planet? Right now we are far from that goal. Around ten million tonnes of carbon dioxide were voluntarily offset in 2005 — less than 1% of the volume and value of transactions in mandatory carbon markets. And 25 billion tonnes of CO₂ entered the atmosphere.

But the voluntary offset market is likely to grow exponentially. The German 3C, which handled the market transactions for FIFA's Green Goal, expects that its project-based transactions, currently 600,000 tonnes of CO₂, will double in 2007. Worldwide, the voluntary market could grow by 40 times by 2010,

"The symbolic value is high. But it's nonsense to believe that a little bit of goodwill is all it takes to solve the problem."

Green activists enlist penguins to save the world



What if threatened species, instead of waiting for extinction, decided to charm humans into saving them? That's the plot of this winter's hit film, *Happy Feet*, in which emperor penguins score a total moratorium on Antarctic fishing thanks to their winsome tap-dancing routines. Environmental groups are using the film, and last year's equally popular documentary *March of the Penguins*, to push their messages — especially the dangers of increasing fishing in the Southern Ocean and climate change.

Mark Stevens, a campaign manager at the National Environmental Trust, sees *Happy Feet* as a way to boost audiences' environmental awareness. "It's a fun, gentle way to educate people about it. Whenever we get a chance, we like to talk about the facts behind the movie." In *Happy Feet*, the penguins' fish shortage is caused by humans. Stevens says that another source of food for some penguins is threatened in real life: "Fishing for krill is expected to increase dramatically in the next few years."

During the height of *March of the Penguins*' popularity, lawyers at another environmental group, the Center for Biological Diversity based in Tucson, Arizona, decided to ask the Environmental Protection Agency to list a dozen penguin species as endangered. They timed the move to the release of *Happy Feet*. "We scrambled to assemble and file the petition while the buzz around the movie was at its peak," says Brendan Cummings, an attorney with the group.

But all this animated advocacy has some US commentators up in arms, criticizing *Happy Feet* as propaganda by stealth. Conservative-leaning Fox News anchor Neil Cavuto called it an "animated *Inconvenient Truth*", referring to former US vice-president Al Gore's documentary on climate change, even though the film doesn't mention global warming directly. Many children's films have had environmental themes in the past few years (see 'Messages in movies').

Meanwhile the US Commission for the Conservation of Antarctic Marine Living Resources, although widely praised for setting sensible fishing quotas, is being pressured by various groups of environmentalists and scientists to consider more cautiously the whole Antarctic ecosys-

COURTESY OF WARNER BROS. ENT.

Gravity probe falters

Two unexpected effects are plaguing a quixotic satellite designed to measure a subtle effect of gravity. But project scientists say they are optimistic that they can still extract a significant finding from their data.

The problems have delayed results from the satellite, known as Gravity Probe-B, until spring 2007, according to project head Francis Everitt, a physicist at Stanford University in California.

Gravity Probe-B is designed to detect a tiny effect of gravity known as frame dragging. According to Einstein's general theory of relativity, a spinning mass such as the Earth drags space-time around it like a mixer dragging batter. Gravity Probe-B hoped to detect the effect with four almost perfectly spherical gyroscopes attached to a telescope. The telescope is trained on a guide star, and the gyros are set spinning. With the

satellite aligned to the star, the drag should tilt the gyros off their axes by a minuscule amount, around 0.000011 degrees in a year.

Gravity Probe-B's journey from conception into orbit was unusually long. It began with modest funding from NASA in 1964, but the high-precision instruments it required took around two decades to develop. The 1986 Challenger disaster further set back the launch of the probe, which was originally scheduled to fly aboard a space shuttle (see *Nature* 426, 380; 2003). After several near-cancellations, the satellite finally lifted off in April 2004. By then it had racked up a price tag of US\$700 million, and another project, Lageos, had become the first to measure the frame-dragging effect (see *Nature* 431, 958; 2004).

After an initial calibration phase, Gravity Probe-B took a year's worth of data,



Gravity effects: wobbles and torques are hampering the interpretation of space data.

completing operations in September 2005. Results were expected by last summer, but the announcement never came.

Nature has learned that this is because two unanticipated effects are clouding the team's frame-dragging result. The first is a slight,

K. STEPHENSON/STANFORD UNIVER., LOCKHEED MARTIN

Messages in movies

Children's films containing environmental messages are nothing new. Here are a few examples.

Ice Age 2: The Meltdown

Pleistocene creatures flee melting ice in this slapstick romance. The melting is caused by local volcanism, but the first line in the film — "This global warming is killing me!" — hints at a climate-change subtext.

Over the Hedge A family of forest critters awake from

hibernation one spring to find half the forest overtaken by suburban development. To find food they must enter human territory to scavenge while avoiding death at the hands of the "freaky pink primates".

March of the Penguins This documentary about emperor penguins lacks the feeling of impending doom present in *Happy Feet*. Tracking their annual journey to breed in the harsh Antarctic terrain, the film focuses on the

life-affirming wonders of nature while steering clear of environmental messages. But packed with the DVD is a short film about research on the effects of climate change on the birds.

Cars Hot-shot racecar Lightning McQueen finds a new friend in Sally Carrera, who laments the construction of the interstate that has cut through the beautiful desert landscape. But the implications of a world of cars for the climate are not explored.

tem, including penguins, when setting fishing restrictions in the Southern Ocean. One major effort is the Antarctic Krill Conservation Project, in which the National Environmental Trust and Pew Charitable Trusts are both involved.

David Ainley, an ornithologist and adviser for ecological consultants H. T. Harvey & Associates, spoke to *Nature* via satellite phone from Ross Island in Antarctica, and understandably hasn't seen *Happy Feet* yet. He's less worried about krill than about Antarctic toothfish (*Dissostichus mawsoni*)



further up the food chain, but adds that the combined effect of climate change and human intervention on the food web are complex and need more research.

Even Ainley's perspective is likely to be more widely disseminated thanks to penguin-mania. He's producing a film on climate change and Adelie penguins, which he hopes to distribute to schools. "Obviously," he says, "we're banking on penguins for the wide distribution of our educational DVD."

Katie McGoldrick and Emma Marris

COURTESY OF WARNER BROS. ENT.

circular wobble created by a combination of tiny imperfections in the gyroscopes and a gradual slowing of their spinning. Gravity Probe-B scientists anticipated the wobble, but their calculations had ruled out the possibility that its period would change. "But it happened," says Everitt. "Nature decides what it wants to do."

The second problem is a torque on the four gyros. The twisting force came about when the satellite's telescope occasionally slipped off the guide star, and it seems to be related to tiny electric fields between various components of the probe. The team didn't discover the effect until after the satellite finished taking data, but are now trying to disentangle it from the result. The problem appeared quite serious during the summer, but Everitt says the team is now working on a "very elegant" solution.

It is unclear how the problems will affect the mission's result. "The Gravity Probe-B team is playing its cards very close to its chest," says Bernard Schutz of the Max-

Planck Institute for Gravitational Physics in Potsdam, Germany. "Nobody outside the project knows whether they can understand and measure the effects well enough to reach their original objective."

That objective was to measure the frame-dragging effect to within about 1%, which would be ten times more accurate than the Lageos result. Clifford Will, a physicist at Washington University in St Louis, who chairs NASA's scientific advisory board for the project and has been briefed by the team, says he is optimistic that there will be some sort of result. "But will it be what they promised NASA? I don't know."

Everitt says that his team will announce a definitive result at the American Physical Society meeting in April 2007. He says he is still worried about whether they can produce a meaningful finding: "I don't think you should be relaxed on \$700 million," he says. "But I wouldn't have agreed to give a talk if I wasn't reasonably confident."

Geoff Brumfiel


**EDISON'S BULBS FAIL TO
LIGHT UP AUCTION**

First all-science collection
sells modestly at Christie's.
www.nature.com/news

CHRISTIE'S

Animal experiments under fire for poor design

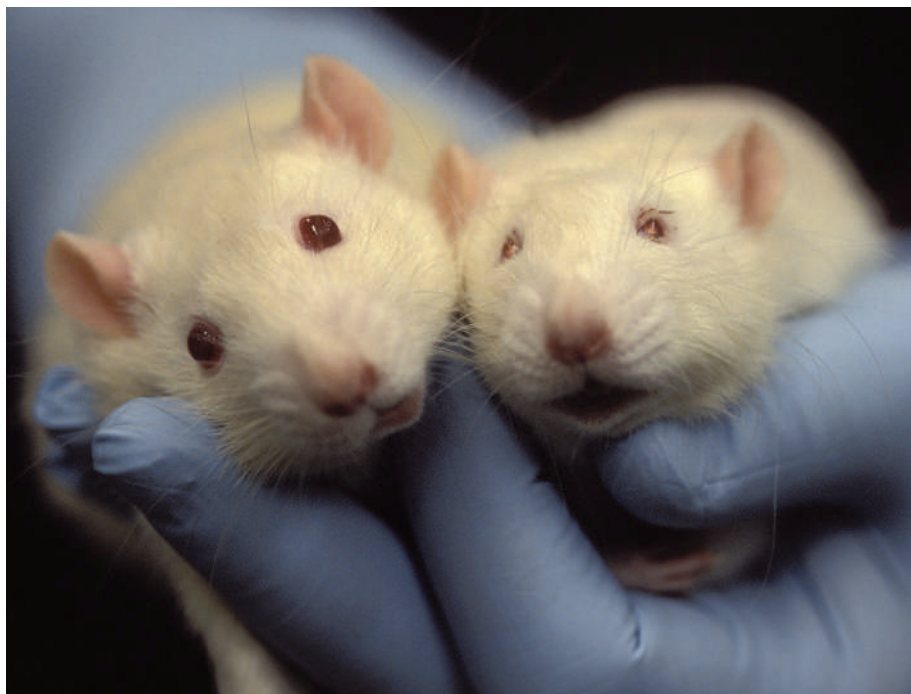
In the contentious world of animal research, one question surfaces time and again: how useful are animal experiments as a way to prepare for trials of medical treatments in humans? The issue is crucial, as public opinion is behind animal research only if it helps develop better drugs. Consequently, scientists defending animal experiments insist they are essential for safe clinical trials, whereas animal-rights activists vehemently maintain that they are useless.

Now a British team has made the first attempt to answer the question in a scientific way, and the result suggests that animal researchers need to raise their game. The team claims that animal experiments are often poorly designed, and so fail to lay the ground properly for subsequent human studies.

The study looked at six treatments that have been evaluated in detail in human trials. The researchers assessed whether animal studies had accurately predicted the outcome of the human work, a task that involved reviewing more than 200 papers. In three of the six cases, the answer was no (P. Perel *et al. Br. Med. J.* doi:10.1136/bmj.39048.407928.BE; 2006).

Ian Roberts, an epidemiologist at the London School of Hygiene and Tropical Medicine and one of the study's authors, says further analysis revealed that poor-quality methodologies and inappropriate models caused the failures. Some studies didn't randomize the allocation of animals to control and treatment groups properly, whereas others showed evidence for publication bias — the tendency not to publish negative results. A third set involved inappropriate models: for example in studies of head injury, treatments were administered just five minutes after rodents were deliberately injured, but humans are typically treated within three hours of accidents.

Roberts would like to see animal researchers learn from clinical trials, where similar problems are now being addressed. He says animal researchers should register experiments before they begin, for example, to ensure the results are reported and to avoid publication bias. He would also like more detailed reporting of experimental procedures, so reviewers can ignore low-quality and potentially biased work. "All the things that happen in clinical trials need



K. KASMAUSKI/CORBIS

Are animals being wasted in badly thought through experiments?

to happen in animal experiments," he says.

Some scientists familiar with both animal and human studies have welcomed the paper, saying it highlights the need for researchers from the two arenas to work more closely on the design of animal experiments. But others have questioned whether a crude comparison

of trials says anything meaningful about the usefulness of animal work.

In many cases, say critics, animal studies are designed to address specific questions not related to a drug's effectiveness, such as how it distributes

around the body, and so cannot be compared to clinical trials. Others point out that applying the standards used in human trials to animal experiments could be prohibitively expensive. For example, initial tests are often run by small groups of researchers. With just one or two scientists, it is impractical to properly blind a study. It is also worth remembering that researchers apply an "intelligent filter" to animal results when designing clinical trials, says Robert Lechler, an immunologist at Kings

College London, and don't expect the results to map straight over to human work.

Roberts counters that his paper looked only at animal studies that were specifically designed to model human disease. And he adds that despite the low cost, small-scale studies are pointless if they do not produce results that people can have confidence in, even if researchers are aware of the limitations. "If you have something cheap and cheerful you have a high probability that the result is wrong," he says. "Just think of the money that is wasted down the line if you follow up that lead."

The results are likely to feature in future campaigns by animal-rights groups as evidence that animal research doesn't benefit medicine. But the authors stress that this would be an error, and that a comparison of just six treatments cannot be used to justify general statements about animal research. "It's not a polemic against animal researchers," says Peter Sandercock, a neurologist at the University of Edinburgh and another author on the study. "But we need to be aware that there are biases in the animal trials."

Jim Giles

"Small-scale studies are pointless if they do not produce results that people have confidence in."

ZOO NEWS SPECIAL

Copycat birth

CC, the world's first cloned cat, has proved her own reproductive credentials by becoming a mother. CC, also known as Copycat, gave birth to three kittens (pictured below) in September, it was revealed this week.

Dog doubles

Meanwhile, Korean embryologists have improved the efficiency of canine cloning. They created three female Afghan hounds from 167 implanted embryos. Last year, Snuppy, the first cloned dog, and a short-lived pup were the only successes from 1,095 embryos.

Dolphin drama

The world's tallest man, Bao Xishun, was called in to save two dolphins that had swallowed plastic shards at an aquarium in Fushun, northeast China. After surgical instruments failed to remove the sharp slivers, Bao used one of his 1.06-metre arms to reach in and extract the chunks from the dolphins' stomachs.

Wimpy seals

Climate change is allowing subordinate males on a remote Scottish island a chance to mate. Higher temperatures and lower rainfall mean that female grey seals forage over a wider range, making it more difficult for the top male to keep an eye on them all.

Rhino romps

Animal-rights activists have slammed plans for a 175-metre sightseeing wheel in Berlin, saying that the bright lights on the structure would disturb the mating habits of endangered rhinos in the nearby zoo.

Sources: BBC, Reuters, Theriogenology

L. WADSWORTH, TEXAS A&M UNIV.

NIH offers free access to wealth of disease data

An unprecedented repository of disease-related data is bringing together information about the genes, health and lifestyles of thousands of subjects studied over many years. The web-based portal will allow any interested investigator to search across multiple epidemiological studies, in the hope of identifying new links to disease.

The 'database of Genotype and Phenotype' (dbGaP) is funded by the US National Institutes of Health (NIH) and was launched earlier this month. So far, it contains data from two big studies that aim to link information about genetic make-up (genotype), physical characteristics including health (phenotype) and lifestyle to particular diseases. These are the 600-subject Age-Related Eye Diseases Study and the 2,573-subject Parkinsonism Study.

Starting next year, the repository will also contain data from the huge Framingham Heart Study, which has followed 14,000 subjects over three generations since 1948. These will be joined by information from the NIH's recently launched Genetic Association Information Network (GAIN), a public-private partnership that genotypes samples from existing studies. Other US and international databases will follow.

It is the first time that some of these big studies, such as the Framingham, have been available to interested parties. And researchers will now be able to mine the vast stores of genetic, phenotypic and study-protocol data simultaneously. "This is really going to change the scope and efficiency of access to the data," says Chris O'Donnell, associate director of the Framingham study.

The repository is part of a larger push towards genome-wide association studies at the NIH. And similar efforts are under way elsewhere. For example, Britain's Wellcome Trust Case Control Consortium began collecting genotype data for such studies in 2005.

Rather than studying only a few genes or gene markers of interest, as researchers have done in the past, these studies try to link features in individuals suffering from certain diseases to variations in a genome-wide set of genetic markers. It is hoped that this will identify new associations, and the huge sample sizes

in the NIH repository should give such studies more power than ever before.

There are technical caveats. If gene variants for a particular disease are very uncommon, or a disease risk comes from hundreds of different genes, then links could still go undetected. Furthermore, data might not have been treated in the same way across different studies, and so may not be strictly comparable. "Controls for schizophrenia might not be as good for diabetes," notes GAIN investigator Pablo Gejman,

"How much of an advantage should be given to researchers who have spent their careers collecting the patient data?"

director of the Center for Genetics in Psychiatry at Northwestern University in Evanston, Illinois. But it is hoped that investigators will gain major insights by looking across studies for genetic variants linked to, for example, blood pressure, which is often measured.

The launch of the dbGaP comes as the NIH wrestles with how best to monitor and disseminate the unprecedented amount of clinical and biological data that studies are now producing. The agency is working on a policy that will govern all the genome-wide association studies that it funds.

There is particular focus on addressing concerns over privacy. Some people worry about the consequences of, say, insurance companies or law-enforcement officials getting hold of the data. The dbGaP's regulations on data access mirror the proposed NIH-wide rules: although study documents, protocols and summary data will be freely available online, researchers will have to apply for access to individual genotype and phenotype results, and these data will lack any information that could identify the subjects.

Also at issue is how much of a publishing advantage should be given to the researchers who may have devoted the best part of their careers to collecting the patient data. The proposed NIH policy calls for a nine-month period during which only the original study investigators can submit a paper for publication based on the data, although all researchers are free to analyse them immediately. In the case of the dbGaP, that period might vary according to the study. Framingham study investigators, for example, will have a 12-month window, according to O'Donnell.

Gene Russo

Fermilab team finds top quark going solo

Physicists at Fermilab near Chicago have obtained solitary top quarks from a mass of particles for the first time — an achievement, they say, that further bolsters the standard model of particle physics.

The top quark — the heaviest of the six quarks — was first observed at the lab's Tevatron accelerator just over a decade ago. But until now, it has been seen only in pairings with other quarks, says Terry Wyatt, a physicist working in the facility's DZero collaboration.

Physicists had long sought a lone top quark because it may provide evidence for heavier quarks. After years of looking, the DZero team announced earlier this month that it had found 60 lone top quarks in a sea of a million billion proton-antiproton collisions. "We're talking about 60 needles in a pretty large haystack," Wyatt says.

The team found no evidence of heavier quarks, but Wyatt says it will continue to search for them.

Chinese river dolphin driven to extinction

Human activity in China's Yangtze river is causing the region's dolphins to go extinct — and more species will follow if fishing is not regulated, conservationists warn.

Scientists who searched the river for the freshwater baiji (*Lipotes vexillifer*), or river dolphin, say it should be declared "functionally extinct" there. That means that even if a few individuals still remain, their numbers will not be sufficient for the species to survive. The dolphin does not live anywhere else — making it the first cetacean reported to be driven to extinction by humans (see *Nature* 440, 1096–1097; 2006).

"There's no hope of saving them," says August Pfluger, chief executive of baiji.org, a Zurich-based foundation that spent six weeks surveying the Yangtze without finding a single baiji. Pfluger adds that other species in the river, such as the sturgeon and the giant salamander, are also threatened.



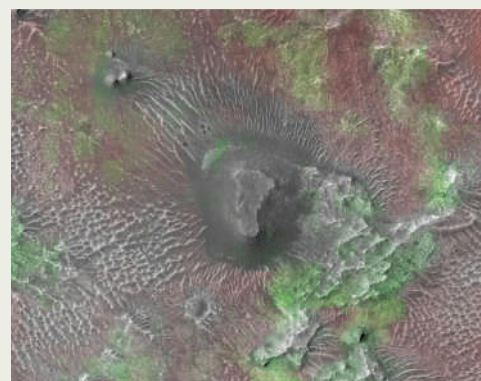
No way back: the freshwater baiji has been described as "functionally extinct".

The red and green planet

This Christmas-coloured image of a martian mesa, courtesy of NASA's Mars Reconnaissance Orbiter, was unveiled at last week's American Geophysical Union meeting in San Francisco. The mineral olivine is shown in red; the green colour represents clay minerals, which might indicate wet environments that could be suitable for life.

The picture, which combines data from two of the spacecraft's instruments, reveals the relationships between these mineral groups. The olivine formed about 3.8 billion years ago; the clay minerals are older still.

"It's important that we've defined that moment in time," says planetary scientist John Mustard of Brown University in Providence, Rhode Island. Now researchers can use that information, he says, to unravel the region's geological history.



NASA/JPL/HUAPL/BROWN UNIV.

Cow-human chimera for stem cells faces UK ban

The British government is on a collision course with some leading stem-cell scientists, after proposing that research on chimaeric embryos be banned.

Teams at Newcastle University and King's College London want to fuse enucleated cow eggs with human sperm in order to extract and study cloned stem-cell lines. An application for the project is due to be considered in the new year by the Human Fertilisation and Embryology Authority (HFEA), the independent body that licenses such research. But proposals for the reform of assisted-reproduction legislation, announced on 15 December, state that such work should be prohibited.

The proposals, which have still to be debated by parliament, will not stop the HFEA from assessing the application in the short-term. And even if passed into law, the ban could be temporary — parliament would be authorized to consider requests to remove the prohibition and allow as yet unspecified research to take place.

Safety of dietary aids comes under rule of law

The US Congress has passed legislation requiring the makers of dietary supplements and non-prescription drugs to inform the government of any serious adverse events resulting from their use within 15 working days of learning of them.

The Dietary Supplement and Nonprescription Drug Consumer Protection Act, sponsored by Senator Orrin Hatch (Republican, Utah), requires supplement-makers to report adverse events, such as hospitalization or death, to the Food and Drug Administration. The requirement will kick in one year after

President George Bush signs the bill into law, which he is expected to do shortly.

Under a 1994 law governing the \$22-billion supplements industry, companies are still exempt from having to prove the safety and efficacy of their products before putting them on the market. But they must now include on the label an address or telephone number for people to report adverse events.

Polonium official danger rating may get upgrade

The poisoning of former Soviet spy Alexander Litvinenko has prompted the International Atomic Energy Agency to consider whether the safety rating of the substance involved — polonium-210 — should be upgraded.

Litvinenko died in London last month after ingesting an unknown amount of polonium-210, which decays by emitting α particles and is fatal in doses of milligrams or below. Distribution of the substance is strictly controlled in Europe and the United States, but tiny sub-lethal amounts are present in products that can be bought without a government licence, such as antistatic brushes used by photographers.

Officials at the atomic energy agency say they are considering whether the danger rating used to compare polonium-210 with other radionuclides used in industry needs to be revised. The substance currently rates 4 out of 5 on a scale in which the most dangerous radiation sources — such as the cobalt used in cancer treatments — are rated at 1.

Correction

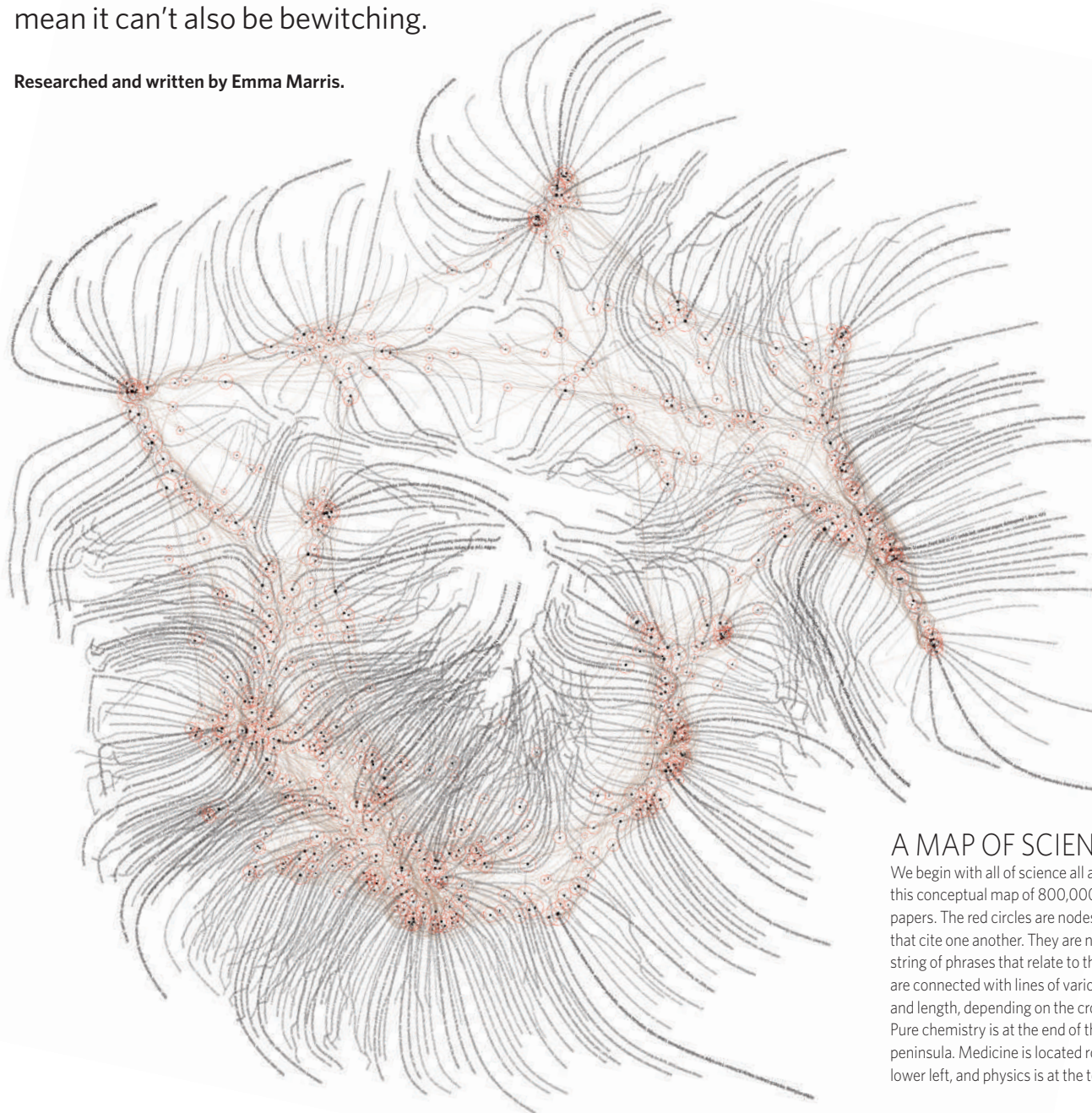
In our News Feature "Mighty mouse" (*Nature* 444, 814–816; 2006) we erroneously stated that 42% of the 50 knockout mouse lines studied showed new phenotypes. The correct figure is 47 lines out of the 50 (94%).

2006 GALLERY

BRILLIANT DISPLAY

From a jewel-like bird, rarer than any diamond, to the delicately poetic swirls generated inside aircraft engines, the pursuit of knowledge turns up its fair share of beauty. This issue, *Nature* wraps up the year with an arresting series of images from 2006. We've divided them into the art of the natural world, planet-scapes both domestic and extraterrestrial, and the splendour of modern technology. Just because something enhances our knowledge doesn't mean it can't also be bewitching.

Researched and written by Emma Marris.



A MAP OF SCIENCE

We begin with all of science all at once, in this conceptual map of 800,000 published papers. The red circles are nodes of papers that cite one another. They are named with a string of phrases that relate to their fields, and are connected with lines of various heaviness and length, depending on the cross-linkages. Pure chemistry is at the end of the right-hand peninsula. Medicine is located roughly at the lower left, and physics is at the top.

K. BOYACK, D. KLAVANS, W. B. PALEY (DATA: THOMPSON ISI, COMMISSIONED: K. BORNER)

YARIGUÍES BRUSH FINCH

Two or three new species of birds are found around the world every year, but not all are as spectacular as this Colombian specimen (*Atlapetes latinuchus yariguierum*), which was announced in June. The area in which it was found is now protected.

B. HUERTAS/AP



WALKING SHARK

A Conservation International expedition in the Indonesian province of Papua recorded 50 new species. Among them was this ambling predator from the genus *Hemiscyllium*, which is just over a metre long and dines on crabs and shrimp while patrolling the sea floor on its fins.

G. ALLEN/CI

K. RASKOFF/MONTEREY PENINSULA COLLEGE



JELLY ROCKET

This 30-centimetre-long siphonophore was captured under the ice in the Arctic Ocean. The spherical objects along its sides are propulsive 'swimming bells'. The bottom — which looks like flame from a rocket — is a mass of tentacles and reproductive organs. The water around it is about 1 °C. This translucent beast won a prize for photographer Kevin Raskoff this July in the BP Kongsberg Underwater Image Competition.



I. STEPHEN

VIRGIN BIRTH

This Komodo dragon was produced by a female kept clear of all males (see page 1021). Parthenogenesis is rare but not unknown in vertebrates, although to date it has mostly been seen in snakes.



SUMATRAN RHINO

This Borneo-based subspecies (*Dicerorhinus sumatrensis harrissoni*) was known about, but never photographed in the wild until June. If we're not careful, this may be our first and last glimpse. There could be as few as a dozen of these animals left.

MOSQUITO IN FLIGHT

Who knew such a pest — and such a killer — could be so beautiful? This *Anopheles stephensi* has just eaten lunch. It was snapped by Hugh Sturrock, who works on fungal pesticides at the University of Edinburgh, UK, and was one of the winners of this year's Wellcome Trust Biomedical Image awards.



H. STURROCK



WWF-MALAYSIA

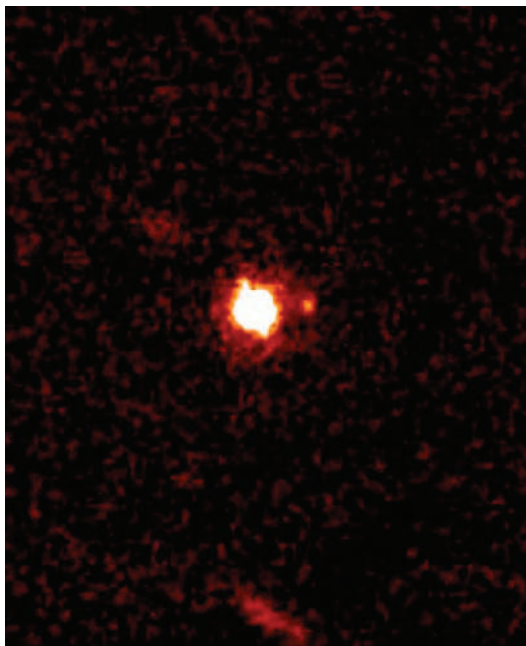


M. KOTTELAT/RAFFLES MUSEUM SINGAPORE

WORLD'S SMALLEST FISH... MAYBE

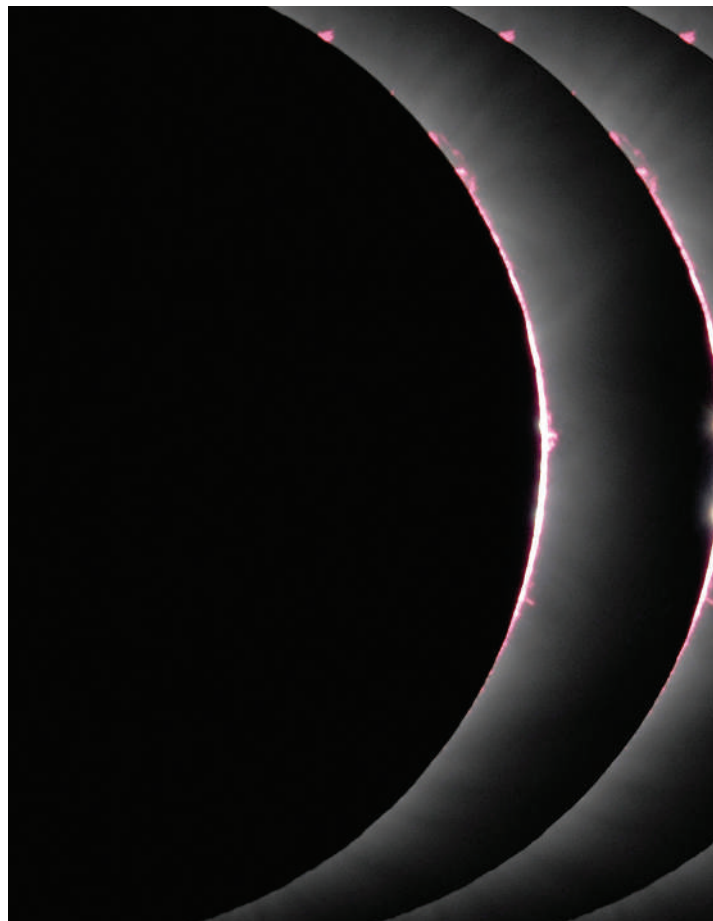
The gauntlets were off this year as *Nature's* online news team stirred up a fight about which is the world's smallest fish. This species, *Paedocypris progenetica*, was identified in January making it the latest contender for the title.

W.M. KECK OBSERV.



◀ THE TENTH PLANET... OR NOT

This body, now known as Eris, was found in 2004, but it was work on its size — it is larger than Pluto — in February that triggered the great planet debate of 2006. That ended in both Eris and Pluto being demoted to the category of 'dwarf planet'.



A MARTIAN LANDSCAPE

The Mars Reconnaissance Orbiter snapped this false-colour picture of the red planet's Victoria Crater. When this photo was taken on 3 October, the rover Opportunity was sitting on the edge of the kilometre-wide crater, although it can't be seen at this scale.

NASA/JPL/UNIV. ARIZONA

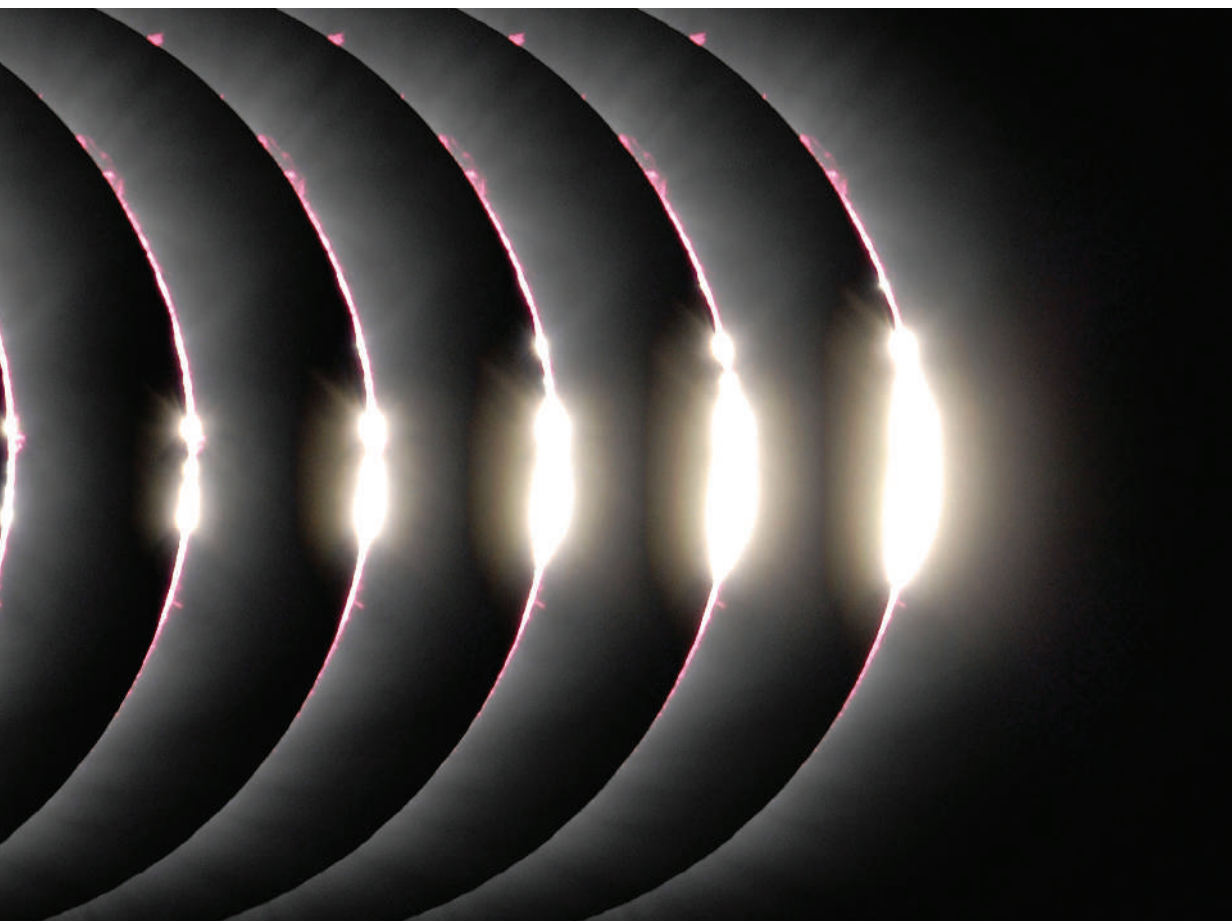


OUR EARTH, EVER MORE FAMILIAR ▶

This is the quarter-of-a-millionth image from the International Space Station, taken on 15 August. The ever-increasing amount of information from space is changing the way we look at ourselves — and encouraging people to abandon the crumpled road maps in their gloveboxes in favour of satellite navigation.



NASA



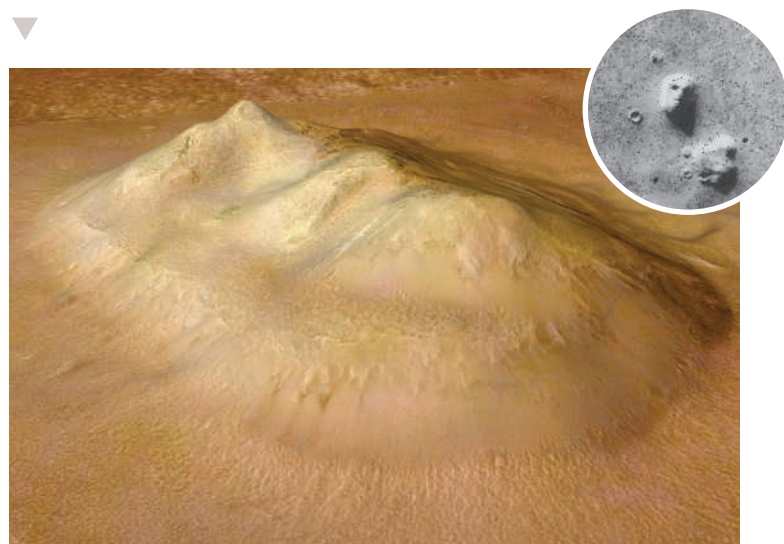
◀ ECLIPSE 2006

A total solar eclipse is not that rare. But it still seems miraculous that the Sun could ever disappear, and images of the event have a unique power. This one was taken on 29 March by Fred Espenak in Libya.

F. ESPENAK, WWW.MIRECLIPSE.COM

A FACE ON MARS NO MORE

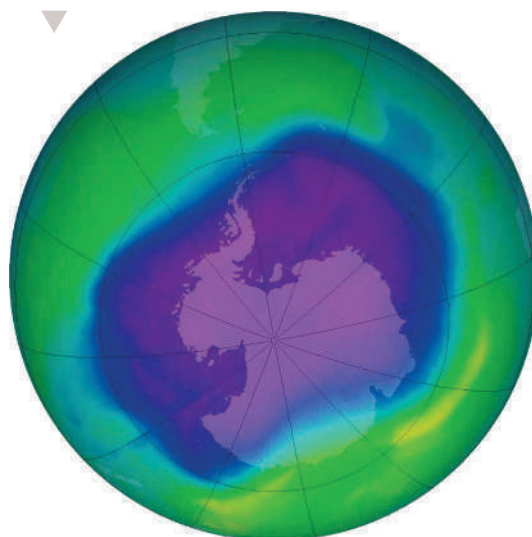
On 21 September, the Mars Express satellite took this image of a cluster of mountains that in a famous 1976 Viking photo (inset) looked eerily like a face. Human beings' innate tendency to see patterns may be at the heart of many scientific discoveries, but we also have it to thank for seeing faces in out-of-focus mountains and crouching leopards in the backyard shrubbery.



ESA/DLR/FU BERLIN (G. NEUKUM), MALIN SPACE SCIENCE SYSTEMS; INSET: NASA/JPL

THE OZONE HOLE IS BACK

The ozone hole is breaking records again. It was as large this September as it has ever been, at 27.5 million square kilometres, thanks to cold Antarctic temperatures. Long term, the hole is still shrinking, but year-on-year variability can mask this gradual process.

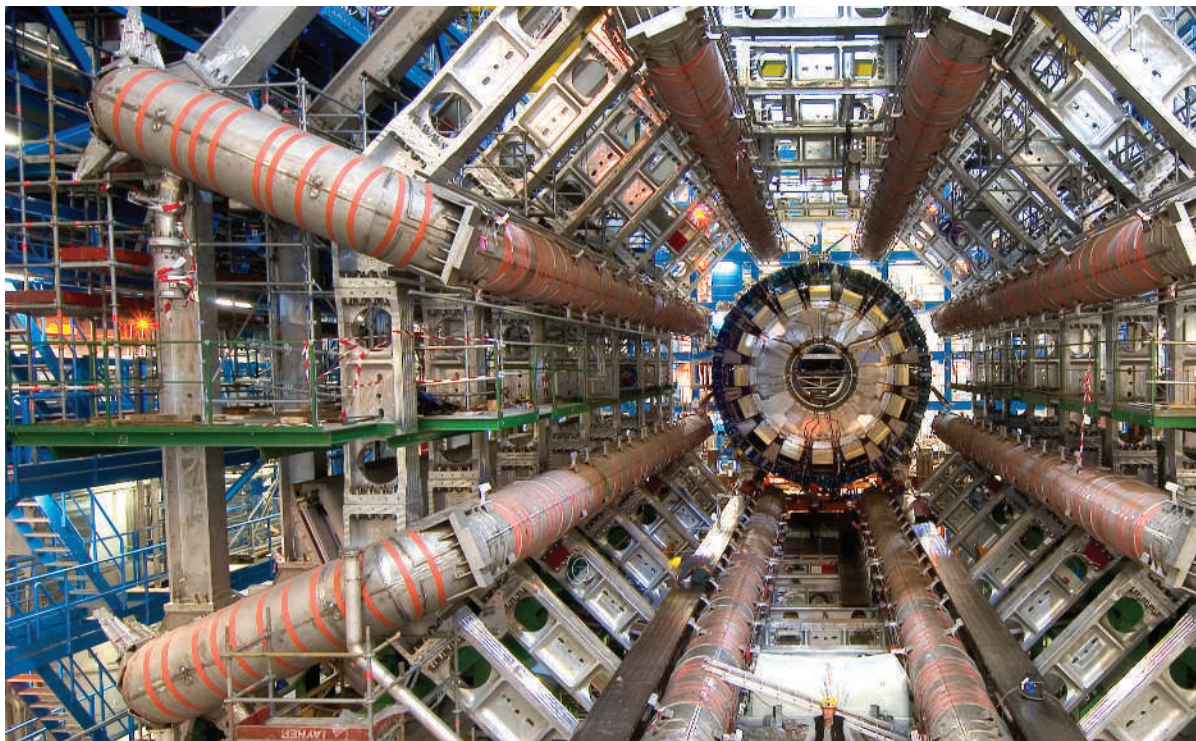


NASA

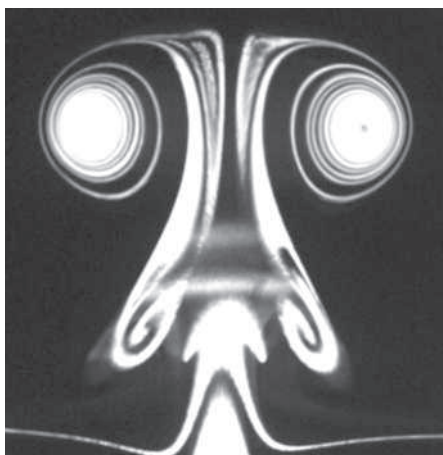
CERN

A CATHEDRAL OF DETECTORS

This is a view down the gullet of the ATLAS experiment at CERN, the particle-physics lab near Geneva. Over the course of this year it has slowly been filled up with all manner of detectors, and work will continue until it's ready to go live sometime in November 2007. Its constructors hope that results from the detectors will allow them to infer the presence of such exotic particles as the Higgs boson.



ONERA

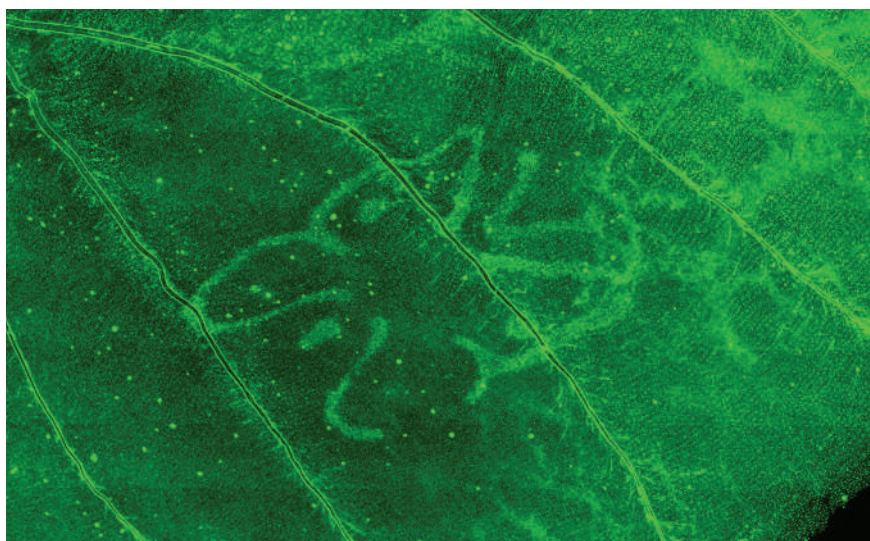


A BUTTERFLY ON A BUTTERFLY

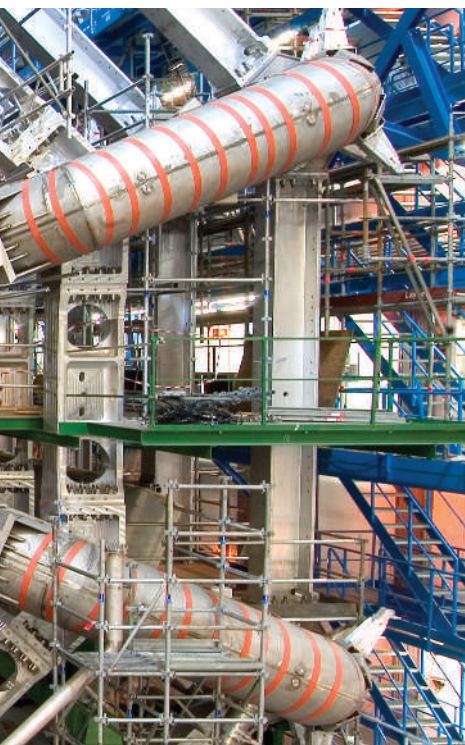
This butterfly, which appeared in November, is actually on a butterfly's wing. The insect was engineered so that its cells expressed a green fluorescent dye when a particular heat-sensitive protein was activated. A laser was then used to induce heat shock and etch out the pattern.

THE ART IN YOUR ENGINE

A vortex of flame swirls inside a combustion chamber. Researchers at the French National Aerospace Research Center (ONERA) caught this shot while examining combustion inside an engine in minute detail. The aim was to make such motors more efficient, but the team has also shown the beauty of their inner workings.

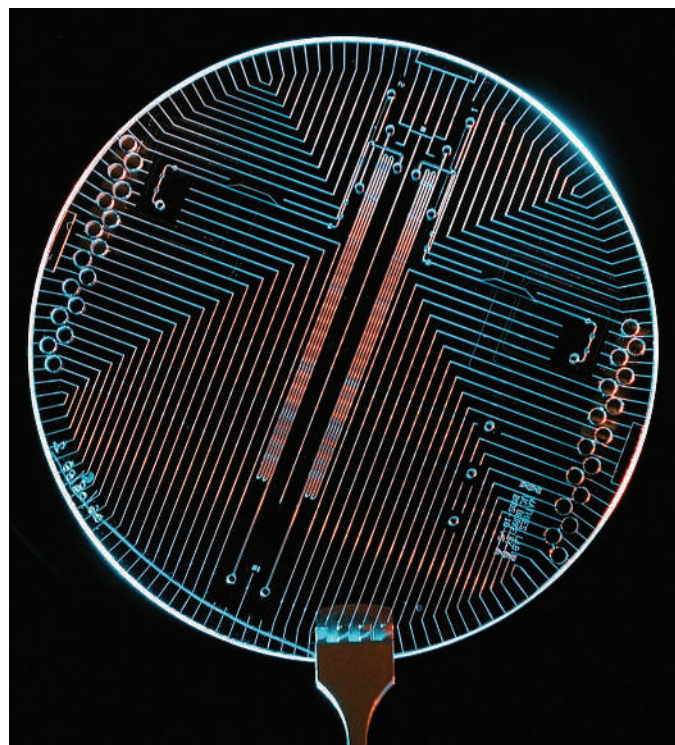


A. MONTEIRO/F. KAMAL/D. RAMOS/UNIV. BUFFALO



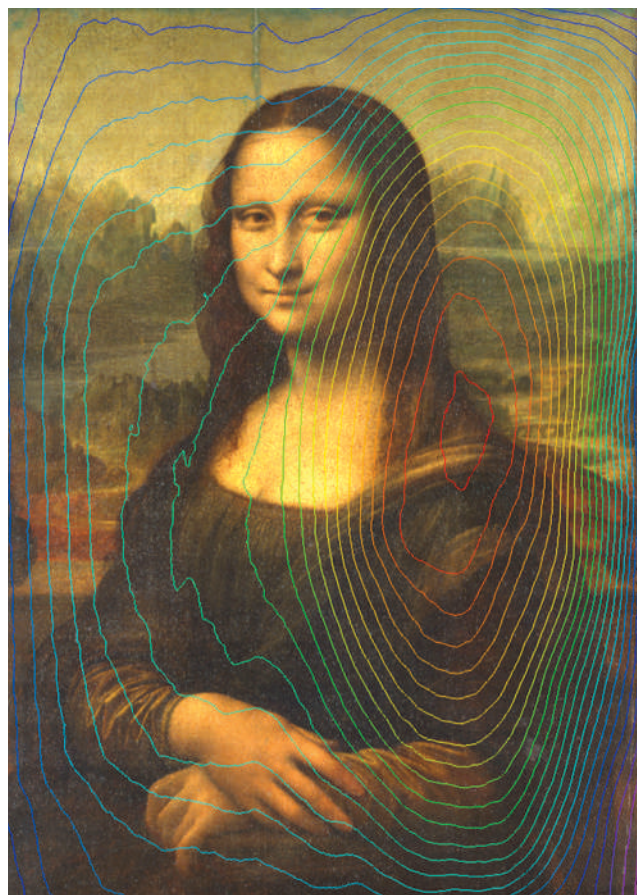
INTRICATE SEQUENCE

Ditch the pipette and do your DNA sequencing at the nanolitre scale with this stylish 100-mm wafer, created by Robert Blazej and his colleagues at the University of California, Berkeley, and unveiled in May. Form follows function in a design that could easily be from an album cover or the floor of a nightclub.



MATHIES LAB UNIV. CALIFORNIA, BERKELEY/PROC. NATL. ACAD. SCI. USA

NATL RES. COUNCIL CANADA



ANATOMY OF A SMILE

In October 2004, the *Mona Lisa* was transported from her home in the Louvre to the Centre for Research and Restoration of the Museums of France in Paris for its first thorough scientific examination. Published in September, this three-dimensional laser scan of the painting's surface shows the slight warping of the wood on which Leonardo painted 500 years ago.

CLING FILM'S CHIC COUSIN

This nanofilm, made by Toyoki Kunitake and his colleagues at RIKEN in Hirosawa, Japan, is a network of organic polymers interlaced with a network of zirconia. Announced in June, the composition of this film makes it ultra-thin (35 nanometres thick), ultra-strong (it can hold 70,000 times its own weight) and ultra-flexible (it can be sucked into a hole 30,000 times smaller than itself).



BUSINESS



IBM's Blue Gene/L uses 130,000 processors.

Better, faster — and easier to use

The Pentagon is sinking millions of dollars into developing the next generation of supercomputers — and plans to let non-military scientists and engineers share the benefits. **Heidi Ledford** reports.

Once upon a time, the fastest supercomputers were top-secret. In the days of the cold war, the best US machines were reserved primarily for the use of spy agencies and designers of nuclear weapons.

But now, the Pentagon's crack research agency is funding the development of machines that, it hopes, will be shared by industry and university scientists, as well as by spies and weapons designers. That way, it figures, more people will write code to harness the computers' massive power — and the cost and effort of developing applications will tumble.

Late last month, the Defense Advanced Research Projects Agency (DARPA) said it would grant almost \$500 million to Cray and IBM to develop machines that will be about ten times faster than the most powerful existing supercomputers — and easier to program.

The award is the third and final phase of DARPA's High Productivity Computing Systems programme, which began in 2002 by supporting research teams at IBM, Cray, Sun Microsystems, Silicon Graphics and Hewlett-Packard. DARPA is managing the programme in partnership with other government agencies, including the Department of Energy, which runs the US nuclear-weapons programme, and the National Security Agency, which spies on radio signals, phone calls and e-mails.

Need for speed

The selection of IBM and Cray was based not only on project design, but also on the commercial viability of the proposed technologies. DARPA has been trying to whet the commercial sector's appetite for supercomputers, and has helped fund a programme within the non-profit Council on Competitiveness to tout the usefulness of the machines.

The two companies will now build prototype

machines, complete with operating systems and software tools, by 2010. The first customers are likely to be governmental agencies. But smaller versions of the machines will also be available to researchers.

Cray has a distinguished history in the field. The world's first genuine supercomputer — the Cray-1 — was released in 1976. It cost \$8.8 million and contained a single central processor, performing up to 80 million floating point operations per second (flops). Today's fastest computer is IBM's Blue Gene/L at the Lawrence Livermore National Laboratory in California, which can perform up to 280 trillion flops and contains more than 130,000 processors.

Processor power

But using vast numbers of processors makes supercomputers tough to program. The more processors you have, the more time they waste just communicating with each other. And although processing speed itself has multiplied over the years, the speed at which they can access computer memory hasn't kept up.

Programmers sometimes try to work around these speed bumps, says Jeffrey Gardner, an astrophysicist and computing specialist at the Pittsburgh Supercomputing Center in Pennsylvania. But he thinks that as more and more processors are added, programming complications will reach the point where scientists simply can't make effective use of the machines. To tackle this problem, both Cray and IBM intend to improve their programming languages so that the computers work more efficiently during lags in communication.

Rick Stevens, a supercomputer specialist at the Argonne National Laboratory in Illinois,

says IBM will focus on optimizing its POWER processors. IBM makes its own chips, he says, giving it more control over their properties.

Cray, on the other hand, buys most of its processors from Advanced Micro Devices. According to Jan Silverman, Cray's vice-president for corporate strategy, it will continue to use these standard processors, which handle one calculation

"Using vast numbers of processors makes supercomputers tough to program."

at a time, but will combine them with Cray-made versions. These will include vector processors, which can perform calculations on multiple pieces of data at once, and multithreaded processors, suitable for database

mining. Cray also plans to optimize performance by using two distinct operating systems — one to perform basic chores such as accessing memory and regulating input and output, and the other to handle data.

Peter Ungaro, Cray's chief executive, hailed the DARPA award as a sign that Cray will return to market leadership. The company has struggled since the early 1990s — Silicon Graphics bought the firm for \$767 million in 1996, only to sell it off again in 2000 for just \$50 million. The DARPA award sent stock prices up by \$2 to almost \$12 — still sharply down from a September 2003 high of more than \$50. Nevertheless, Stevens is optimistic about Cray's outlook, saying that several recent contracts bode well.

But Cray still derives most of its research and development money from government grants, as the market for supercomputers isn't large enough to support their high development costs. "The military has always been at the forefront of supercomputing," says Robert Deupree, a physicist at St Mary's University in Nova Scotia, Canada, "in part because they needed it, but also because they had the money to be able to invest in it."

How Africa learned to love the cow

The development of lactose tolerance in sub-Saharan Africa is a fascinating tale of genetic convergence, reports **Erika Check**.

The Dinka people of southern Sudan, it is said, have 400 different words to refer to the cattle that they prize above all other things. The Maasai, who live in Kenya and northern Tanzania, have traditionally believed that all cattle on Earth were given to them by the gods, and value those in their possession accordingly. The Zulu, Xhosa and Swazi in South Africa devote themselves to their strikingly patterned Nguni cattle, whose hides are now prized by high-end interior designers.

And in all these pastoralist or semi-pastoralist groups, which rely on herding for survival, people drink milk — which is something of a puzzle. Most adult humans (those of European stock are largely an exception) find the sugar in milk, called lactose, indigestible. The gene for lactase, the enzyme that breaks lactose down into the more digestible forms of glucose and galactose, is normally switched off as children are weaned. Without lactase, lactose is of little use to a milk drinker; but it is still a valid food for some stomach bacteria, which can have unsettling and unpleasant results. Thus, most adults tend not to drink much milk.

Now, genetic detective work is showing how, in parts of Africa, evolution found a way round this problem — just as it did a few thousand years earlier in northern Europe. The results should help clarify the origins and spread of cattle rearing in Africa, and provide a textbook illustration of the ways in which the same social innovation can write its consequences into the human genome in different times and places.

A drinking game

Sarah Tishkoff of the University of Maryland, College Park, first heard the story of 'lactase persistence' — the ability to continue making lactase throughout adult life — in an introductory anthropology course she took at college in the late 1980s. In the 1960s, anthropologists first started noticing that groups of humans that raise cattle also tended to be the same ones that could drink milk as adults¹. In the 1970s, a genetic basis for lactose intolerance was established, although the nature of the mutation has remained a mystery². Drinking milk thus came to be seen as a metabolic, not merely cultural, proclivity. People whose ancestors had herded cattle had evolved to make use of their milk.

In places where lactase persistence is rare, the milk that is drunk is frequently sour — its lactose content lowered by hungry bacteria.

But it was not until 1997, when Tishkoff was doing postdoctoral research with Andy Clark at Pennsylvania State University, that she began thinking about lactase persistence as a focus for her own work. Tishkoff visited geneticist Trevor Jenkins at the University of Witwatersrand in Johannesburg, South Africa, to help him study an extensive collection of DNA samples from southern Africans. She began to wonder whether she could trace the genetic origin and history of lactase persistence by studying the DNA of Africa's cattle herders. If so, it would be a neat piece of genetics. And it would also be of interest to anthropologists interested in cattle domestication itself. Had the ancestral aurochs been domesticated just once, in the middle East, and the practice then spread to Europe and Africa? Or had Africans learned to domesticate cattle on their own?

Tishkoff designed a project to collect DNA samples from 43 ethnic groups in three East African countries and to correlate the data with various physiological characteristics relevant



to taste perception, disease susceptibility and other things — including lactose metabolism. It was an ambitious plan at best — a foolhardy undertaking at worst. Few researchers, if any, had ever collected such a large, diverse set of genetic samples from Africa. Many of the groups she wanted to study were in remote locations, isolated by choice or by circumstance. And although she started to apply for permission to research in Tanzania in 1999, she didn't get permission to work there until 2001.

Living on the edge

Tishkoff set out with a 30-year-old Land Rover full of cases of energy bars and camping gear. Over the next three years, on and off, she and her students roamed around Tanzania, Kenya and Sudan, collecting data in some dangerous as well as out of the way places — such as parts of Kenya where carjacking was rampant and everyone carried semi-automatic weapons. The team also had a close call once, on a journey near Lake Eyasi in Tanzania, when a bus rounded a corner and smashed head-on into the Land Rover; miraculously, no one was hurt. Tishkoff credits her students — especially Jibril Hirbo and Alessia Ranciaro, who collected samples in the most dangerous parts of Kenya — with persistence and courage. “Those guys literally risked their lives for this work,” Tishkoff says.

At most stops, Tishkoff's group hired a translator to explain the purpose of her research, get permission from local leaders, and recruit

people for their research. In all, they took samples from 470 people. To test for lactase, the team members stirred powdered lactose into cups of water, gave it to each person, and took a series of timed blood samples. That told the researchers how well each person could digest lactose, as well as providing samples of the individual's DNA.

In 2002, while the work was under way, a team of Finnish researchers reported that it had found a genetic mutation that seemed to cause lactase persistence in North Europeans³. A small change in an ‘enhancer’ region upstream of the lactase gene seemed to keep the gene from being switched off after infancy. All of the 137 lactase-persistent Finns studied had the mutation in question, and studies of other populations showed that the frequency of the mutation matched that of the lactase persistence. But two years later, another team reported that the Finns' genetic variant was not found in East Africans, even though some are lactase persistent⁴.

Tishkoff's team showed why this was. Ranciaro sequenced DNA from her subjects and noticed several mutations close to the lactase gene. A collaborator at the Wellcome Trust Sanger Institute in Cambridge, UK, Panos Deloukas, checked for the mutations in the full set of 470 people, and Tishkoff and a

postdoctoral fellow working with her, Floyd Reed, noticed that one of these mutations was tightly linked to lactase persistence in Tanzanians and Kenyans. The team also found that two other mutations were associated with lactase persistence in people in northern Kenya and Sudan, though not as strongly⁵. It was immediately clear that these mutations had arisen independently from the one found in Finland.

“Until the geneticists contributed to the data, we thought about evolution as happening very slowly and gradually.”

— Diane Gifford-Gonzalez

Quick start

The DNA also showed evidence that the mutation seen in the Tanzanians had spread very quickly. DNA accumulates small, random imperfections over the generations, and yet the stretches of DNA surrounding the lactase persistence mutation were identical in most of those who

had it. This uniformity shows that the gene evolved recently and spread rapidly — which in turn means that it must have conferred an advantage strongly selected for by evolution. Statistical models of Tishkoff's data suggest that the mutation arose between 3,000 and 7,000 years ago — a blink of an eye in evolutionary time.

Tishkoff and Reed conclude that lactase persistence bears one of the strongest signatures of positive selection ever observed in the human genome. Mutations that favour

SVEN TORFINN/PANOS



malaria resistance, such as the sickle-cell gene and an inability to make the enzyme glucose-6-phosphate dehydrogenase, or G6PD, are also strongly favoured and spread rapidly. (They seem to have got going at about the same time, too, and are also related to domestication — malaria is thought to have first become a major problem for Africans when they started to live in settlements.) But the positive selection for lactase persistence seems even stronger, perhaps because the costs of the mutation are less severe than those for malaria. Tishkoff and Reed suspect that the advantage might go beyond the extra calories that could be gained from the lactose. Lactase persistence might also have allowed people to stay alive during times of drought, when those benefiting from the mutation would have been able to drink milk without the risk of diarrhoea, which exacerbates dehydration.

The study, and subsequent follow-up, should help to elucidate the origins of East African

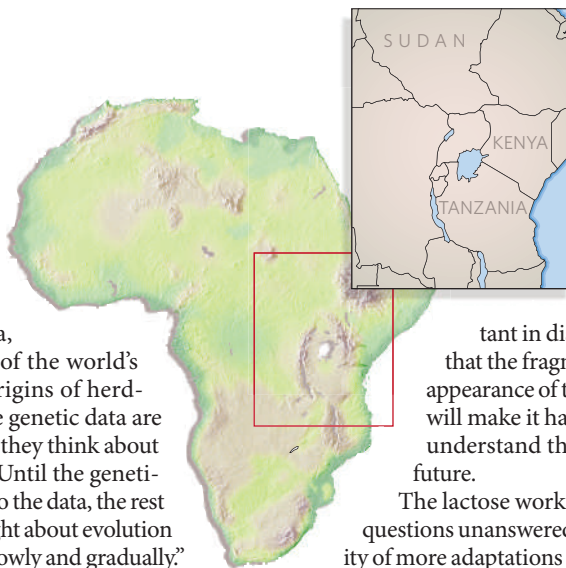
cultures and traditions. Anthropologists such as Diane Gifford-Gonzalez from the University of California, Santa Cruz, one of the world's experts on the origins of herding, say that these genetic data are changing the way they think about human history. "Until the geneticists contributed to the data, the rest of us always thought about evolution happening very slowly and gradually."

Herding along evolution

And the genes provide data even in places, such as East Africa, where the archaeological data are poor. "That is why this article is so interesting to people like me," says Gifford-Gonzalez. "This gives us a totally independent line of evidence about the origins of dairy-

ing, and gives us a much better way of homing in on when these major nutritional transitions occurred." The human genetic data might thus in some ways complement the genetic data on cattle, which in the decade since Tishkoff began thinking about the issue have come down in favour of multiple domestications, including one in Africa.

Tishkoff's work also highlights the incredible genetic diversity in African people, a diversity that as yet has been studied very little. For example, the African DNA samples in the Haplotype Map⁶ — the catalogue of genetic diversity published last year — come from only one ethnic group: the Yoruba of West Africa. But, traditionally, the Yoruba don't herd cattle, and Tishkoff didn't discover any mutations for lactase persistence in them (although in some populations of West African pastoralists, fewer people have the 'European' lactase persistence mutation). Such blindspots could be a problem, because scientists hunting for the genetic causes of diseases often rely on HapMap data. "There's such a huge amount of genetic diversity in Africa that we're clearly going to have to look at all the ethnic groups in different regions to find all the variants,"



Tishkoff says. "Otherwise, we could be completely missing things that are impor-

tant in disease." She worries that the fragmentation and disappearance of traditional cultures will make it harder to access and understand that diversity in the future.

The lactose work still leaves a lot of questions unanswered, and the possibility of more adaptations still to be found. A particularly intriguing question surrounds the Hadza of Tanzania, who show a surprisingly high level of lactase persistence despite having very little to do with cattle. One possibility is that, though they are now mainly hunter-gatherers, their ancestors might have been pastoralists. Another is that lactase may offer some other benefits besides the ability to drink milk — perhaps aiding in the digestion of some specific local foods.

A textbook case

But if there are still questions, there's also a significant achievement. Scientists have to date found little genetic evidence for convergent evolution in people: "This is the best example of convergent evolution in humans that I've ever seen," says Joel Hirschhorn, a geneticist and paediatrician at Children's Hospital Boston, Massachusetts. "Lactase persistence has always been a textbook example of selection, and now it'll be a textbook example in a totally different way."

Convergent evolution is not unknown in humans; lighter skin colour seems to have evolved independently in Europe and Asia, and a range of different malaria adaptations are known. But lactase persistence offers a particularly simple and tractable example: there's a single gene involved, with different mutations in different parts of the world having similar effects. The challenge now is to learn from this textbook example how to spot more subtle convergences that have been forced on human biology by shared experiences and cultural innovations — or that are still under way today. ■

Erika Check is a senior reporter for Nature in San Francisco.

1. Swallow, D. M. *Annu. Rev. Genet.* **37**, 197–219 (2003).
2. Sahi, T. *Scand. J. Gastroenterol. Suppl.* 1–73 (1974).
3. Enattah, N. S. et al. *Nature Genet.* **30**, 233–237 (2002).
4. Mulcare, C. A. et al. *Am. J. Hum. Genet.* **74**, 1102–1110 (2004).
5. Tishkoff, S. A. et al. *Nature Genet.* doi:10.1038/ng1946 (2006).
6. The International HapMap Consortium *Nature* **437**, 1299–1320 (2005).



Cattle call: the ability to drink milk is a useful trait with a variety of origins.

T. VOETEN/PANOS; G. PIROZZI/PANOS; C. HUGHES/PANOS



CREATING THE PERFECT WAVE

Kerry Black travelled the world in search of the best surf spots. Then he decided to build them himself — on land. **Mark Schrope** meets the maverick oceanographer.

Lori Wilson Park, near Cape Canaveral on Florida's east coast, is not known as one of the world's best surf spots. On a typical weekday approaching sunset, a handful of surfers might be found catching some good, but not great, waves. Florida's finest surfing doesn't usually occur until a tropical storm or other suitable weather system passes the right distance offshore.

Surfers here may grumble at times, but they accept they are at the mercy of the ocean's fickle moods and the beaches' idiosyncrasies. But surprisingly enough, that may be about to change. Soon, surfers at Lori Wilson Park could become the first ever to have a choice between two new surf options: catching natural waves that have been dramatically improved by an artificial reef, or driving an hour inland to ride great artificial waves at a newly developed surf park.

When that happens, they can thank Kerry Black, a physical oceanographer and lifelong surfer who has spent the past decade trying to override the ocean's whims. He plans to install offshore artificial reefs at places to better control waves and, in the process, improve surfing, create new habitats for marine life and prevent beach erosion. At the same time, Black and his New Zealand-based company, ASR, are work-

ing with US partners to construct the first wave pool purpose-built for surfing — which would generate phenomenal waves rivalling those at the world's best natural surf spots.

Breaking technology

Black seems well qualified for both tasks. Born in Melbourne, Australia, he began surfing when he was 14 and eventually trained as an oceanographer. For decades, his research focused on topics such as larval dispersal along reefs and sand transport along beaches. But in 1995 he accepted a professorship at Waikato University in Hamilton, New Zealand — a position that, he says, offers a great deal of autonomy “as long as you don't break any obvious rules”.

To the university's surprise, Black decided to shift his research focus. “About two months in,” he remembers, “I was literally wandering through the grounds wondering what I wanted to do when I thought: let's see if we can understand surf breaks. And that's when we set up the programme.”

His wave-research programme included travelling to 44 of the world's best surfing spots — from Bingin in Bali, to Malibu in California — to gather sonar data on the topography of the reefs there. Black was flooded with applica-

tions from students who wanted to be involved, but faculty response ranged from shocked to cynical to confrontational. In time, the university community came to accept the research, says Black, but the larger academic world was not as malleable. “I took the responsibility to actually explain to every student that if this is your degree course, you could be studying for a profession that does not exist,” he says.

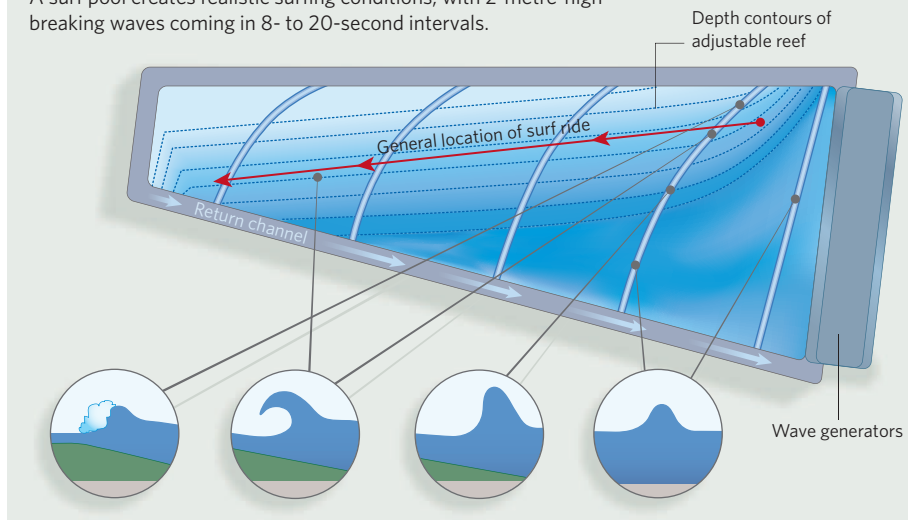
Today, that profession has to some degree emerged. After seven years at Waikato, Black left to form his company with some of his graduate students. ASR is based in the surf town of Raglan, made famous in the 1966 documentary *The Endless Summer*. The company's main interests are in designing and building surf pools and artificial reefs; dozens of the latter are in various stages of development in nations such as India, South Africa and the United Kingdom. And in August, the company sponsored the biennial International Surfing Reef Conference, a three-day event that drew some 50 researchers to Black's own surfing resort on Lombok, Indonesia. There, for the first time, many of the presentations were able to go beyond theory and into case studies of reefs that have actually been built.

“If you add up all the surfing breaks in the

N. BOTHMA/EPA/CORBIS

MAKING WAVES

A surf pool creates realistic surfing conditions, with 2-metre-high breaking waves coming in 8- to 20-second intervals.



world and count how many you would call classic surf breaks, you realize it's a very small fraction," says Black. This suggests that a nearly magical combination of conditions must align to create great natural waves. Black hopes to capitalize on his knowledge of the world's classic surf spots to build artificial reefs that create the same sort of surfing experience.

Wave theory

The exact details are proprietary, but in the simplest terms, the important factors are the slope of the reef on its ocean side — which helps determine when and how a wave will break on the landward side — and the angle of the reef in relation to incoming swells, known as the 'peel angle'. For good surfing, it is essential that waves do not break all at once but peel — or break in such a way that surfers can ride in front of the crest on an open wave face, or even inside the breaking portion, known as the barrel.

Other critical wave characteristics are size and power; not surprisingly, the bigger and more powerful the wave, the better the surfing. Artificial reefs can control these factors, to some extent, by concentrating the waves' energy. The surfer's ideal wave usually requires a light wind blowing out from shore, to groom the wave face and make it smoother — but reef designers can't do much about that.

Artificial reefs are not as scenic as natural ones; at ASR, they are made of huge sand-filled bags, up to 50 metres long, made from synthetic fibre sheets. Every location has different wind and ocean swell conditions, so each reef must be designed individually for that spot. Reef designers must also meet specific wave goals: for instance, whether to maximize the number of surfable days per year or the quality of waves when the very best swells come through. Reefs can also be tailored for interme-

diate or advanced surfers. "There is a gigantic juggling exercise that goes on in trying to balance all these different effects," says Black.

Another factor to juggle is whether the artificial reefs can help protect the beach from erosion. Florida's Lori Wilson Park, like many beaches worldwide, has to be repeatedly replenished by sand dredged from offshore, a process called 'nourishment'. "It is basically buying you time, because you're putting the beach back to where it used to be," says Robert Dalrymple, a coastal engineer at the Johns Hopkins University in Baltimore, Maryland. "One thing it doesn't do is cure the original cause of the beach erosion. If it eroded before, it's going to erode again."

Such nourishment projects can protect beachfront properties and restore vanishing beaches, but also have high financial and ecological costs. Black and his colleagues think artificial reefs could hold a solution.

"I was wondering what I wanted to study when I thought: let's see if we can understand surf breaks."
— Kerry Black

Artificial reefs could work in some instances, says Dalrymple, as anything placed offshore that reduces the wave energy hitting a beach could potentially reduce erosion. But he warns that the solution isn't necessarily simple. "You have to be really careful because you can also put a structure offshore that makes the erosion problem worse," he says. For instance, an artificial reef could capture sand that would have otherwise been transported to — and deposited on — another section of beach.

The first test of one of Black's designs came at the Narrowneck Reef on Australia's Gold Coast, a massive 60,000-cubic-metre structure completed in 2000. The project, funded mainly to protect the beach from erosion, also included dumping nourishment sand in areas up and down the beach, either side of the reef, to balance erosion that might occur as sand shifted and the beach adjusted to the new structure.

"So far things are going great," says John McGrath, an engineer with the Gold Coast City Council who oversees the monitoring work. The beach inside the reef has widened and nearby erosion has proceeded gradually at expected rates, although McGrath says problems could conceivably still arise given that the period since construction has seen fewer than normal severe storms.

Against the current

Very little is really known about how artificial reefs can affect erosion, says Muthukumar Narayanaswamy, a graduate student at Johns Hopkins who works with Dalrymple. He recently completed a white paper on the potential of artificial reefs to prevent erosion. For instance, a small artificial reef off El Segundo, California — not built by ASR — settled into the sea bed so deeply that it neither improved surf nor reduced erosion. "A lot more work needs to be done before saying that these structures are the cure-all for coastal erosion," he says.

Narayanaswamy's study was commissioned by the Surfrider Foundation, based in San Clemente, California. Chad Nelsen, the group's environmental director, is also cautious; he worries that artificial reefs could create a false sense of security that might promote even more coastal development than has already occurred in the United States. Nelsen thinks proposals for artificial reefs should be carefully scrutinized. Black, he says, "is a salesman and also a talented scientist, but he's motivated to build reefs".

At Narrowneck, the artificial reef hasn't been as good at enhancing surf as it has at preventing beach erosion. Waves have improved somewhat, but not as much as intended. Although designed by ASR, the company was not involved in its construction, which was handled by local engineers. The bags that make up the reef were dropped from a barge, and the reef ended up not being as high as planned. This kept the barge from grounding itself on the bags, but also reduced the reef's impact on incoming swells.

But Narrowneck has worked well ecologically. The reef is covered with seaweed that supports marine creatures such as worms, crabs, lobsters and a wealth of baitfish. Corals cannot attach themselves to the geotextile material, but there are few nearshore reefs in the area, making it biologically valuable, says Steve Smith, a marine biologist at the University of New England in Coffs Harbor, Australia. Narrowneck is also popular with snorkellers.

Elsewhere, ASR has had more luck with achieving both good waves and good ecology at its 6,000-cubic-metre Mount Reef in Mount Maunganui, New Zealand. To get around the barge depth problem, Black's team designed a system for connecting the bags using seatbelt material. Workers dragged a web of empty bags into place using anchors in the sea floor, then filled them with sand using dredging equipment. "This is the first reef we've had control



M. SCHROPE

Brain wave: physical oceanographer Kerry Black wants to create an artificial reef and ocean-sized swell in a wave pool for surfers in Florida.

over all the way," says Black. "It is really proof of all parts of the concept."

Mount Reef saw its first significant swell in October. "To see some waves breaking as we knew they would, that's a pretty good feeling," says David Neilson, executive director of the Mount Reef Trust, which raised money for the project. Sand has also built up between the reef and the beach, and crayfish have colonized the reef, along with mussels, crabs and snapper.

Surfers' paradise

But even the world's best artificial reef relies on storms, winds and tides to create great waves. To get around that problem, Black's team set out to design a new kind of wave pool specifically for surfers. The first is under construction in Florida.

There are countless wave pools around the world already, mainly for lazy, family fun in water parks. Although some of the pools produce surfable waves, none are specifically designed for surfing. Black realized that a few relatively simple changes to basic designs could dramatically improve the situation. Wave pools typically get wider as you move away from the wave machine, and they use fresh water. To improve surfing, ASR's design calls for narrowing pools to concentrate energy and salt water to improve buoyancy.

More critically, wave pools don't contain reefs. So ASR designed a patented structure called the Versareef, made of steel triangles

overlain with rubber, to sculpt pool waves to satisfy even the most advanced surfer. The shape of the Versareef can be shifted on demand to allow different configurations — thus producing waves that mimic those of the world's best surf breaks.

The first Versareef pool is a beginner's model that will produce waves about 1 metre high, at the Ron Jon Surf Park in Orlando, Florida. If this is successful, a larger pool, for expert waves up to 2 metres high, will be built soon after that. A number of other projects, including in England and California, are on hold until the first pool is successfully completed. "No one will sign until this one works," says Black.

In August, the pool was filled and the first waves created, but the prototype Versareef was unable to withstand the pressure of the waves. On a recent visit to the test pool, despite the setback, Black spoke confidently of the prospects for success. "We know it's going to work," he says. The reef structure has now been beefed up substantially, and the first successful waves were produced at the beginning of December.

For Florida surfers, the pools clearly can't be completed soon enough; thousands are already on the waiting list for park membership. Black envisages that one day the pools

will host unprecedented surf contests where competitors will be able to ride identical waves, which will enable more accurate judging. And he hopes the pools could even be used for wave research, as they will be larger than most wave tanks that are currently available to scientists. Dalrymple says the pools could be a valuable resource if they are not too expensive for scientists' research budgets.

Back in Cocoa Beach, Florida, surfers may also see better surfing options soon. John Hearin, an engineer with NASA, is working with Black and others to push through plans for an artificial reef that Hearin designed for his master's degree. The plan, most likely for Lori Wilson Park, is still in its early stages, but local

officials and business have been supportive.

So, with natural waves enhanced by artificial reefs and unbelievable artificial waves available on demand, will surfers be willing to give up their famous global quest for good waves? Absolutely not, says Black. He views the pools and artificial reefs as a great addition to surfing, but not as replacements for the real thing. Surf beaches are his favourite place in the world to go, he says. "I really need these places for my own health to get away and just surf."

Mark Schrope is a freelance writer, and regular surfer, in Florida.

"To see some waves breaking as we knew they would, that's a pretty good feeling."
— David Neilson

FREAKS OF NATURE?

Ultraendurance racers torture their bodies and minds to achieve near-impossible physical feats. Is it an exceptional genetic make-up or the vestiges of human evolution? **Helen Pearson** reports.

Trek 125 kilometres, and cycle 250 more. Kayak 131, rappel through canyons for another 97, and swim 13 in churning whitewater. Throw in some horseback riding and rock climbing; spread it all over six days in the blistering Utah heat; and never stop to sleep.

That's the punishing formula for the annual Primal Quest adventure race, considered one of the most extreme tests of human endurance. Of the 90 four-person teams that competed this summer, only 28 finished the course. "It was more like adventure torture," says Michael Tobin of the winning team.

Yet a growing number of athletes are opting to push the outer limits of their body's abilities. If a marathon or Ironman triathlon isn't enough, athletes can enter 'extreme ultraendurance' events, which spread the agony over several days (see 'The toughest races on Earth'). And alongside the racers — figuratively and sometimes literally — are exercise physiologists.

Researchers see these radical events as a testbed for ideas on how some people manage to perform physical feats at which others can only marvel. Some scientists argue that ultraendurance athletes have special genes or physiology that allow them to perform beyond ordinary limits. Others claim that these athletes are little different physically from fit people such as marathon runners — but have something unique about their brains that allows them to keep exercising when the body screams 'stop'. Either way, many find it hard to resist the opportunity to study human physiology as it is stretched towards breaking point. "Scientists are naturally drawn to the extreme," says Euan Ashley, a cardiologist at Stanford University in California who has studied the genetics of endurance racers.

Yet much of the research is still in its infancy. Most of the focus in exercise physiology remains on prestigious professional and Olympic sports, not the protracted and gruelling competitions of amateurs. In addition, ultraendurance is difficult to study in the lab because few subjects are willing to run on a treadmill for 24 hours or more.

But some key differences between ultraendurance athletes and other athletes are beginning to emerge. For instance, short races, such as 10-kilometre competitions, are usually won by athletes in their 20s. But ultraendurance races tend to be won by people in their early to mid-30s — because the body takes a longer lifetime of training to achieve the toughness of muscle and efficiency of metabolism that such a race demands, experts say. Ultraendurance winners also are good at performing for sustained periods of time at 60–70% of their maximum output. The strongest conventional athletes, by contrast, are those who use oxygen most efficiently when working at close to 100% of their maximum.

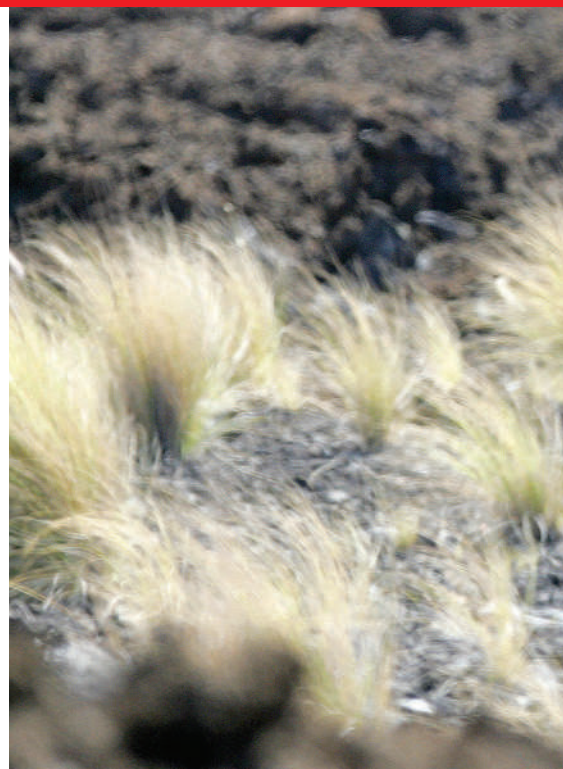
Pushing the limits

Additional insights are starting to come from studies led by Mikael Mattson and Jonas Enqvist, researchers at the Karolinska Institute in Stockholm, Sweden. Their team had nine world-class racers run, cycle and kayak on lab machines nonstop for 24 hours while the researchers recorded various measures of heart function and metabolism. The group is still analysing the data, but one preliminary result suggests that ultraendurance exercise makes the energy-generating mitochondria in the athletes' muscles more efficient at using fat rather than glucose as fuel.

Fat generates more energy than glucose per kilogram, but the body can't normally burn fat when exercising very hard. The bodies of endurance racers may have found a way to go faster using fat as fuel, saving stores of glucose for later on in the race. "It's one reason these athletes are physiologically different from other athletes," Mattson says. He calculates that the athletes burned around 20,000 kilocalories (84,000 kilojoules) during the day of exercise — an amount almost impossible to replenish by eating.

What is not clear from Mattson's work is whether the ultraendurance athletes were born with special physiological abilities or whether they gained them through training. But genes probably play at least some part, according to

"The human body is well-adapted to long-distance running, as an evolutionary hangover from our hunting and scavenging days."



Energetic mutation: are Ironman world champions, such as 2005 winner Faris Al Sultan above, endowed with endurance genes?

studies of more than 400 participants in the South African Ironman competition.

A team led by Malcolm Collins, a molecular biologist at the University of Cape Town, has examined the gene that makes angiotensin-converting enzyme (ACE), a peptide that helps narrow blood vessels. In a separate study, one form of this gene had already been found to be associated with endurance performance in mountaineers and army recruits¹. More recently, Collins' group found that more than 77% of the fastest South African-born Ironman triathletes carried one or two copies of this endurance variant, compared with 67% of non-athletes². The effect didn't hold for non-South African athletes, but it triggered Collins to start looking at other potential gene variants that could affect Ironman performance.

In particular, he is now scrutinizing two genes that work along the same biochemical pathways as ACE. Variants of these genes affect the blood flow and metabolic efficiency of muscles: one makes a receptor for a peptide called bradykinin, and the other makes an enzyme called nitric oxide synthase 3. The researchers found that athletes with a particular combination of 'fitness' variants of these two genes finished the Ironman in 12 hours 30 minutes, on average, compared with 13 hours 4 minutes among those who carried the 'slow' variants³.

Tim Noakes, a sports physiologist who works with Collins at the University of Cape Town, argues that the bodies of ultraendurance athletes are actually little different from those of other sports people. The brain, he says, is the oft-overlooked organ that sets ultraracers



kilometres at a time helped our ancestors survive, he says. Fuelled by plentiful water, energy bars and yet more training, that body can complete the 90 or more kilometres of an ultramarathon.

If this is true, then many of those who study ultraendurance racers have also embraced their evolutionary past. Paul Laursen, for instance, was kicked out of kinesiology studies at Simon Fraser University in British Columbia, Canada, after becoming so absorbed in triathlon competitions that he flunked his first year. Now he has a remarkably lean 10.6% body fat, a 48-beat resting heart rate and, in early December, swallowed a miniature thermometer and ran an Ironman in Western Australia as the subject of his own study into muscle damage. He ran a worse time than any of his previous Ironmans — but says the experiment was a success.

Helen Pearson is a reporter for *Nature* in New York.

1. Montgomery, H. E. *et al. Nature* **393**, 221–222 (1998).
2. Collins, M. *et al. Med. Sci. Sports Exerc.* **36**, 1314–1320 (2004).
3. Saunders, C. J. *et al. Hum. Mol. Gen.* **15**, 979–987 (2006).
4. Baden, D. A., McClean, T. L., Tucker, R., Noakes, T. D. & St Clair Gibson, A. *Br. J. Sports Med.* **39**, 742–746 (2005).
5. Bramble, D. M. & Lieberman, D. E. *Nature* **432**, 345–352 (2004).

apart. These people “are mental freaks”, he says, not physiological ones.

Conventional thinking in exercise physiology holds that the muscles, lungs and cardiovascular system set the limits of performance. But the brain coordinates biochemical signals from these peripheral organs, such as those indicating fatigue and tissue damage, and will unconsciously stop us before we wreck our bodies. Perhaps, Noakes suggests, ultraendurance athletes have brains that allow them to override warning signals that would bring others to their knees.

Noakes is now examining what conscious and unconscious information the brain uses to set the ideal exercise pace. In one recent study,

he and his colleagues showed that athletes use oxygen more economically and feel less exertion if they know the length of time they will be running, compared with those deceived about the time but who run at exactly the same pace⁴.

But even couch potatoes may have something of the endurance racer in them. Daniel Lieberman, a biological anthropologist at Harvard University in Cambridge, Massachusetts, argues that the human body is well-adapted to long-distance running, as an evolutionary hangover from our hunting and scavenging days⁵. Ultraendurance racers “are able to be freaks because evolution has enabled us”, he says. A body capable of jogging tens of

“The brain is the oft-overlooked organ that sets ultraracers apart — they are mental freaks, not physiological ones.”

E. KREUTZ/NEWSPORT/CORBIS

The toughest races on Earth

Robyn Benincasa, a San Diego firefighter and successful ultraendurance athlete, remembers her most gruelling moment. In the Ecuadorian Andes in 1998, she had been racing for two days and nights without sleep, and faced climbing a volcano more than 6,000 metres in altitude. “My nail beds were blue, my lips were blue, my whole body was blue,” she recalls. “I was on my hands and knees in the snow, crying.” Somehow — she does not remember how — she made it to the summit, and her team eventually won.

Sound like a good time? People voluntarily participate in such races, looking to push their bodies to the limit. Here’s a look at some of the most challenging:

Race Across America

A non-stop bike race that spans the United States. Deprived of sleep, competitors cover more distance than the Tour de France in about 40% of the time.

Badwater Ultramarathon

A 215-kilometre foot race, billed as the world’s toughest, that starts 85 metres below sea level in California’s Death Valley and goes to around 2,500 metres at Mount Whitney. Temperatures can reach 55 °C. The current male record holder took 24 hours, 36 minutes to finish.

Hawaii Ironman Triathlon

The world championship of Ironmans spans a 3.9-kilometre swim, a 180-kilometre bike



ride and a 42-kilometre run. Participants win a spot in qualifying Ironman races around the world.

Marathon des Sables

A week-long, 240-kilometre ultramarathon in the Moroccan desert. Competitors carry their own food and gear.

Primal Quest Expedition Adventure Race

The biggest, and most publicized, adventure race. Teams of four trek, cycle, paddle, navigate, climb, canyon and otherwise cover roughly 700 kilometres in a quest for a prize worth US\$250,000 in 2006.

H.P.

R. OLIVER/EPA/CORBIS

Life: perhaps we should take the porridge theory with a pinch of salt

SIR — In his Book Review of *The Goldilocks Enigma* ("Life in the universal porridge" *Nature* **444**, 423–424; 2006), Jim Al-Khalili discusses Paul Davies' answers to the question "Why is the Universe just right for life?"

The question relies upon the widely held assumption that life, in particular human life, is a unique form of matter requiring a special explanation for its existence.

As a biologist, I agree that life is a fascinating phenomenon, and humans may well be among the most complex entities in the Universe.

However, all past and present life-forms on Earth have been the products of one particular evolutionary process, which was driven by natural selection and which operates on the basis of specific genetic mutations and selective environments. If these mutations and environments had been different, the products of this evolutionary process would also have differed. Indeed, the organisms we know of, including our own species, might never have evolved at all.

Thus, our existence does not require a special category of explanation, beyond natural selection. To think so can be seen as a form of chauvinism — seductive but misguided — that elevates the specific aggregates of matter on this planet, and their historically conditioned features, to an inappropriately lofty status.

This argument also applies to SETI, the search for extraterrestrial intelligence, which is based on the same tempting assumption. Life, consciousness and intelligence, as we know them, are highly unlikely to be found elsewhere in the Universe, nor are they likely to be inevitable consequences of any sufficiently 'advanced' darwinian system. Rather, they are specific outcomes of one particular process of evolution, occurring on one of the billions of planets in this Universe. (If the Multiverse exists, there could be other universes almost identical to ours, and therefore there could be other life-forms like us, but presumably we would not be able to communicate with them.)

These conclusions will be disappointing to many — as they are, sometimes, to me. But, as Richard Dawkins commented about people who have been disturbed by his book *The Selfish Gene*, in his Darwin@LSE public lecture in London this year, "If something is true, no amount of wishful thinking can undo it".

Rather than leaving these matters, as Davies suggests, to "philosophers and priests", or to physicists, perhaps we should listen to the biologists as well.

Then some of these questions might be considered less troubling, and less important, than they currently seem to be.

Peter Wigley

22 Chatham Road,
Keswick, South Australia 5035, Australia

Life: porridge would be just right for each universe

SIR — Jim Al-Khalili's Book Review of Paul Davies' *The Goldilocks Enigma* ("Life in the universal porridge" *Nature* **444**, 423–424; 2006) explains with great clarity why the anthropic principle is an incomplete answer to the great question of 'Life, the Universe and everything', and why the concept of a Multiverse is more satisfactory.

However, there is an alternative — even if we assume that the only way to make a different universe is to change a few fundamental constants.

Arguments in favour of fine-tuning typically show that some key ingredient of our current Universe, such as atoms or stars, becomes unstable if some physical constant is changed by a relatively small amount and therefore cannot exist in a universe with different constants. But in such circumstances, what is interesting is not the instability of some particular state, but what the system does instead. If conventional atoms or stars become unstable, what other organized forms of matter might arise? In particular, which values of the physical constants permit structures complex enough to resemble intelligent life?

It is not easy to answer this question. At the moment, we cannot rigorously deduce the structure of the helium atom from basic physics, let alone that of a living organism. Our understanding of atomic structure depends on various ad hoc simplifications, justified by comparison with observation. This approach is not available for universes with different constants, so we have no real idea how they might behave.

If — as experience with simpler dynamical systems suggests — large regions of universe-space permit complex structures, then our Universe actually has a reasonable chance of falling into such a region, and the complex structures that arise will necessarily be consistent with the laws of that universe.

The real message of Goldilocks is not that Baby Bear's porridge was 'just right'. It was just right for Goldilocks — but cold porridge was also just right for Mummy Bear and hot porridge was just right for Daddy Bear. We are the product of our own universal porridge, and it will always be just right for us.

Ian Stewart

Mathematics Institute, University of Warwick,
Coventry CV4 7AL, UK

Biography of Crick aims to inspire a wider audience

SIR — We do hope that your readers are not put off reading Matt Ridley's biography *Francis Crick: Discoverer of the Genetic Code* (HarperCollins, 2006) by Horace Judson's review (*Nature* **443**, 917–918; 2006). Ridley's book, which, as one of a series of biographies, had to conform in length to a standard, makes no pretence to be the academic biography that Judson was apparently expecting. He should know (as Ridley does, and makes clear in his acknowledgements) that the 'definitive' biography of Crick is being written by Bob Olby. Ridley set out to write about 'the greatest biologist of the twentieth century' for a wide audience.

Judson comments, for example, that insufficient detail was given about Crick and Watson's interactions with the Wilkins and Franklin group. But these interactions have been considered by historians many times, including by Judson himself in his book *The Eighth Day of Creation* (Simon & Schuster, 1979), and could only have been condensed in this brief account.

Ridley conveys, in a way that is both clear and affectionate, Crick's unique mix of fun and genius, as well as the highlights of his scientific achievements. Judson calls Ridley a 'journalist'. That's an arguable description, but in a world where the communication of science to a general audience is so essential, we wish there were more 'journalists' like him. His biography of Crick will make many eager to read more about this inspiring man.

Michael Ashburner*, **Mark Bretscher†**,

Peter A. Lawrence‡

*Department of Genetics,

†MRC Laboratory of Molecular Biology,

‡Department of Zoology,
University of Cambridge, Cambridge, UK

Shellfish view of omega-3 and sustainable fisheries

SIR — The mackerel pictured in A. K. Duttaroy's review of Susan Allport's book *Queen of Fats* (*Nature* **444**, 425; 2006) are indeed a source of omega-3s, but others are just as good and more sustainable. Early humans evolved eating fish, probably intertidal shellfish, while living a shoreline existence in Africa. They didn't have boats to trawl for pelagic oily fish such as mackerel, nor did they have hooks and lines. Now, as then, intertidal herbivorous shellfish such as mussels and clams can help people reach a healthy 1:1 balance of omega-3 and omega-6 fats in their diets, instead of the current 1:17.

Anthony Robson

Institute of Environmental Sustainability, Swansea
University, Singleton Park, Swansea SA2 8PP, UK

COMMENTARY

Measures for measures

Are some ways of measuring scientific quality better than others? Sune Lehmann, Andrew D. Jackson and Benny E. Lautrup analyse the reliability of commonly used methods for comparing citation records.

Although quantifying the quality of individual scientists is difficult, the general view is that it is better to publish more than less and that the citation count of a paper (relative to citation habits in its field) is a useful measure of its quality. How citation counts are weighed and analysed in practice becomes important as publication records are increasingly used in funding, appointment and promotion decisions. Typically, a scientist's full citation record is summarized by simpler measures, such as average citations per paper, or the recently proposed Hirsch index¹, which is ever more being used as an indicator of scientific quality². Despite their growing importance, there have been few attempts to discover which of the popular citation measures is best and whether any such measure is statistically reliable.

Measures of citation quality are of value only if they can be assigned to individual authors with high confidence. Previous bibliometric studies² have compared different measures of scientific quality, but just because two measures agree does not mean that either one is accurate or reliable. We will argue that some citation-based measures can provide useful information given data of sufficient quality, but others fail to meet minimum acceptable standards. This should concern every working scientist.

Unfair discrimination

Because citation practice differs markedly between disciplines and subfields, a homogeneous set of authors is essential for any statistical analysis of citations. Here we use data from the theory section of the SPIRES database in high-energy physics, which has the requisite homogeneity³. Within this database, the probability that a paper will receive k citations falls slowly with increasing k and is described by a power-law distribution, a/k^γ with $\gamma \approx 2.8$, for large k . This long-tailed distribution has a number of consequences. About 50% of all papers have two or fewer citations; the average number of citations is 12.6. The top 4.3% of papers produces 50% of all citations whereas the bottom 50% of papers yields just 2.1% of all citations. Measuring an

author's mean or median citation count per paper probe different aspects of their full citation record: which is better? Fortunately, this question can be posed in a way that yields a statistically compelling answer.

The purpose of comparing citation records is to discriminate between scientists. An author's citation record is a list of the number of citations of each of the author's publications. Until reduced to a single number, this list cannot provide a means of ranking scientists. But whatever the intrinsic merits of the chosen number, it will be of no practical use unless the uncertainty in assigning it to individual scientists is small. From this perspective, the 'best' measure will be that which minimizes uncertainty in the values assigned and hence maximizes discrimination

between individuals. We analyse three measures of author quality: mean number of citations per paper, number of papers published per year, and the Hirsch index. A scientist is said to have Hirsch index h if h of their total, N , papers have at least h citations each, and the remaining $(N-h)$ papers have fewer than h citations¹. For this study, we adopt Hirsch's assumption that h divided by N "should provide a useful yardstick". To calibrate our results, we also consider an obviously meaningless measure; we rank authors alphabetically by name.

We start with one of the three

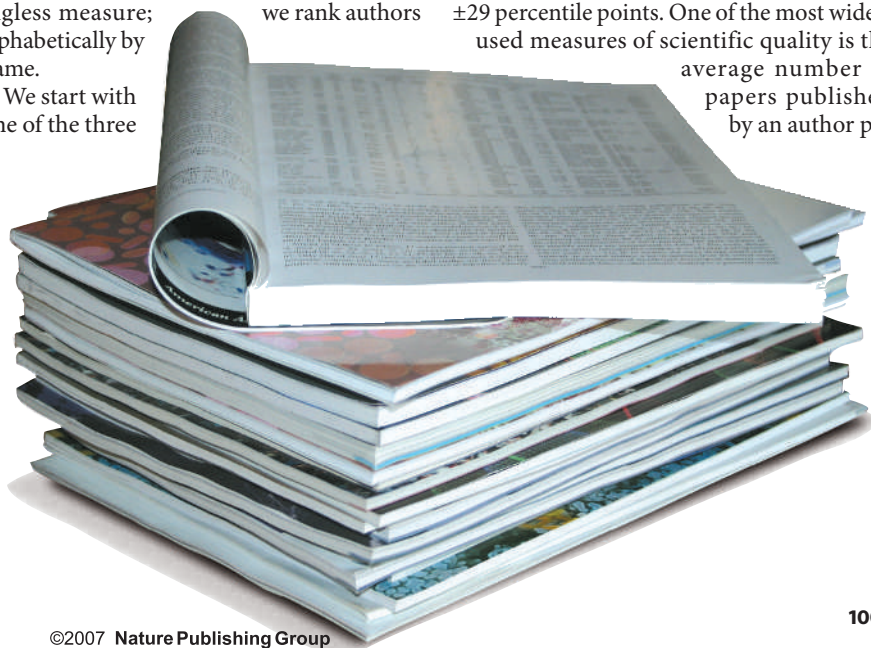
measures we had chosen and sort the SPIRES authors into decile bins. We use the full citation records for all authors in a given bin, n , to calculate the conditional probability that a paper written by an author in bin n will have k citations. From these conditional probabilities, we use Bayes' theorem to determine the average probability that an author initially assigned to bin n should instead be assigned to bin m . (To do this, we calculate the probability that the full publication record of each author in bin n was drawn, at random, on the conditional probability appropriate for bin m ; see Supplementary Information.) Because the m assignment is based on an author's full citation record, it is more reliable than the n assignment. This process is repeated for each decile bin and for each measure considered.

Quality testing

A perfect measure of author quality would place all weight in the diagonal entries of a plot of m versus n (Fig. 1, overleaf). The better the measure, the more weight will be found in the diagonal boxes. Figure 1 reveals that both accuracy and certainty are sensitive to the choice of indicator.

An alphabetical ranking of authors contains no information regarding scientific quality, and so every author is assigned to every decile with equal probability (Fig. 1a). The resulting root-mean-square (rms) uncertainty in author assignment thus has the maximum value of ± 29 percentile points. One of the most widely used measures of scientific quality is the average number of papers published by an author per

"There have been few attempts to discover which of the popular citation measures is best and whether any are statistically reliable."



A. MACDONALD

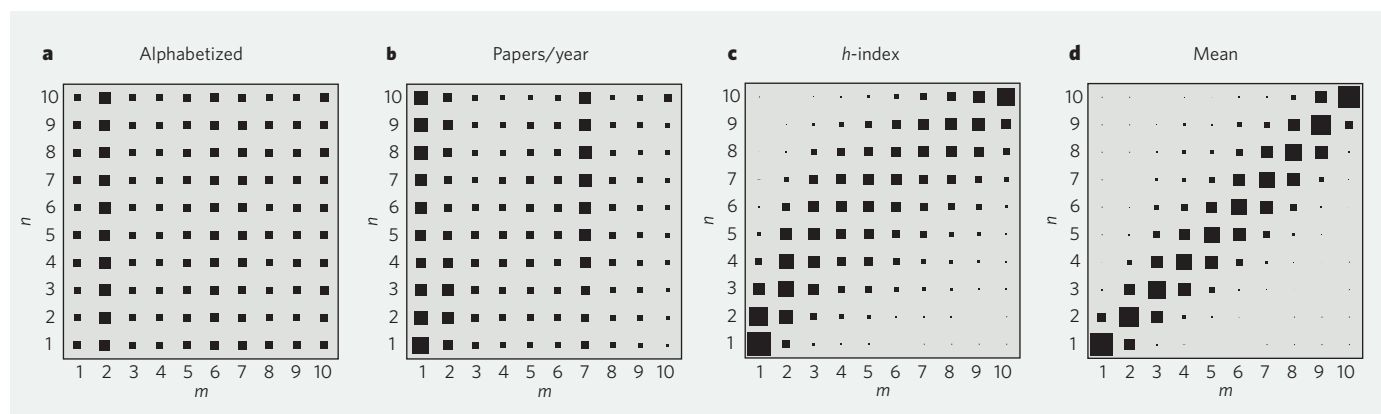


Figure 1 | The probabilities for four different measures. **a–d**, Each horizontal row, indexed by n , shows the average probabilities that authors initially assigned to a given decile bin n are predicted to lie in a different decile bin m . The probabilities are proportional to the areas of the corresponding squares.

year^{1,4}. This measure has a similar *rms* variation to alphabetization (Fig. 1b). Publication frequency would be more useful if all papers were cited equally but, as noted above, this is not the case. The best that can be said of publication frequency is that it measures industry rather than ability.

Impact factors are widely used to introduce a citation measure into calculations of publication frequency. The impact factor for each journal, as defined by Thomson Scientific/ISI⁵, is the average number of citations acquired during the past two years for papers published over the same period. But weighting each paper by the journal's current impact factor is unlikely to improve the situation, especially when estimating scientific quality across an author's entire career. The impact factor for reputable journals is determined by a small fraction of highly cited papers, so the citation rate for individual papers is largely uncorrelated to the impact factor of the journal in which it was published⁶. The widespread use of publication frequency — with or without an impact factor — is disturbing and requires further study.

Word of caution

Hirsch's h -index attempts to strike a balance between productivity and quality and to avoid the heavy weight that power-law distributions place on a relatively small number of highly cited papers. Hirsch's measure is obtained by ranking papers in order of decreasing citations with paper i having $C(i)$ citations and solving the equation $h = C(h)$. This is the simplest version of $h = AC(h)^K$. Hirsch's choice of $A = K = 1$ is unsupported by any data. Nevertheless, Fig. 1 indicates that this measure does better than publication frequency, because the h -index depends on the entire citation record.

Hirsch's measure overestimates the initial n -assignments by some 8 percentile points as indicated by higher densities above the diagonal (Fig. 1c). Moreover, the *rms* uncertainty in the assignment of h is ± 16 percentile points, which is only a factor of two better than alphabetization. Although capturing certain aspects of quality, Hirsch's index cannot make

decile assignments with confidence.

Compared with the h -index, the mean number of citations per paper is a superior indicator of scientific quality, in terms of both accuracy and precision. The average assignment of each n -bin is in error by 1.8 percentile points with an associated *rms* uncertainty of ± 9 . Similar calculations based on authors' median citation give an accuracy of 1.5 and an uncertainty of only ± 7 percentile points, suggesting that the median copes better with long-tailed distributions.

Simple scaling arguments⁴ show that the *rms* uncertainty for any measure decreases rapidly (exponentially) as the total number of papers increases. Thus, for example, no more than 50 papers are required to assign a typical author to deciles 2–3 or 8–9 with 90% confidence when using the mean citation rate as a measure. Fewer papers suffice for deciles 1 and 10. Any attempt to assess the quality of authors using substantially fewer publications must be treated with caution.

Data access

The methods used here are not specific to high-energy physics. Given suitably homogeneous data sets, they can be applied to any scientific field and permit a meaningful (probabilistic) comparison of scientists working in different fields by assuming the equality of scientists in the same percentile of their respective peer groups. Similarly, probabilities can be combined to make meaningful quality assignments to authors with publications in several disjointed subfields.

There are strong indications that an author's initial publications are drawn on the same probability distribution as their remaining papers⁷. Therefore, with sufficient numbers of publications to draw meaningful conclusions (50 or more) the mean or median citation counts can be a useful factor in the academic appointment process.

Unfortunately, the potential benefits of careful citation analyses are overshadowed by their harmful misuse. Institutions have a misguided sense of the fairness of decisions reached by algorithm, and unable to measure what they want to maximize (quality), institutions will maximize what they can measure. Decisions will continue to be made using measures of quality that either ignore citation data entirely (such as frequency of publication) or rely on data sets of insufficient quality.

Access to the full citation distribution for an entire subfield is essential to our analysis. Existing databases such as the ISI can therefore actively help to improve the situation by compiling field-specific homogeneous data sets similar to what we have generated

for SPIRES. This would allow institutions and scientists alike to evaluate the quality of any citation record using all available information. For their part, scientists should insist that their institutions disclose their uses of citation data, making both data and the methods used for data analysis available for scrutiny. In the meantime, we shall have to continue to do things the old-fashioned way and actually read the papers.

Sune Lehmann is at the Department of Informatics and Mathematical Modeling, Technical University of Denmark, DK-2800, Lyngby, Denmark.

Andrew D. Jackson and Benny E. Lautrup are at The Niels Bohr Institute, Blegdamsvej 17, DK-2100, Copenhagen, Denmark.

1. Hirsch, J. E. *Proc. Natl Acad. Sci. USA* **102**, 16569 (2005).
2. van Raan, A. F. J. *Scientometrics* **67**, 491 (2006).
3. Lehmann, S., Lautrup, B. E. & Jackson, A. D. *Phys. Rev. E* **68**, 026113 (2003).
4. van Raan, A. F. J. *J. Am. Soc. Inf. Sci.* **57**, 408 (2005).
5. Thomson Scientific/ISI <http://www.isinet.com/>
6. Seglen, P. O. *J. Am. Soc. Inf. Sci.* **45**, 1 (1994).
7. Lehmann, S. Thesis, The Niels Bohr Institute (2003).

Supplementary information accompanies this Commentary on Nature's website.

BOOKS & ARTS

The ambiguity is the essence

When we consider the Universe, are we trying to impose order on a meaningless jumble?

The Human Touch: Our Part in the Creation of a Universe

by Michael Frayn

Faber & Faber: 2006. 505 pp. £20

To be published in the US in February by Metropolitan Books

Alan Lightman

"What do you think of tastes — do they exist without the mind or no?" asks Philonous in philosopher George Berkeley's 1713 treatise *Dialogues Between Hylas and Philonous*. The materialist Hylas replies: "Can any man in his sense doubt whether sugar is sweet, or wormwood bitter?" The antimaterialist Philonous (representing Berkeley himself) then asks: "Is a sweet taste a particular kind of pleasure or pleasant sensation?" and "is not bitterness some kind of uneasiness or pain?" Hylas concurs and Philonous concludes: "If, therefore, sugar and wormwood are unthinking corporeal substances existing without the mind, how can sweetness and bitterness, that is, pleasure and pain, agree to them?"

The ancient philosophical question, of course, is this: does the Universe have any meaningful existence beyond our perception of it?

The celebrated playwright Michael Frayn is the latest to tackle this problem, in his new book *The Human Touch*. Breathtaking in its sweep and audacity, this is a huge sprawl of a book. Ranging from quantum physics to neuroscience, philosophy, linguistics and literature, Frayn brings every discipline he can muster to argue that Philonous is right: the world has no order or meaning beside what we give to it. According to Frayn, the laws of nature, cause-and-effect relationships, even time and space, are concepts largely manufactured in our minds. "Logic, it turns out, is a theoretical construction, something we have built ourselves," he writes.

To Frayn, the physical Universe is a meaningless jumble, filled with uncertainties and unpredictabilities, governed by chance and often random events. Like actors in a Greek tragedy, we are driven to find order and meaning. We develop language, we name things, we try to analyse our own consciousness and 'I-ness', we wrestle to rationalize the decisions we make, we create categories of truth and falsehood, we invent ideas of causality and logic, and we propose laws of nature — but all these attempts at order are ultimately doomed to result in ambiguity. For not only is the external Universe a



K. WOTHE/MINDEN PICTURES/GETTY IMAGES

Mind over matter: we impose our ideas on the world when we try to make sense of what we see.

complex mess, so too are our minds, subject as they are to intricacies beyond any rational understanding.

Above all else, Frayn rails against determinism, against a materialist and causal explanation of the world, from a man falling in love with a woman to Earth's orbit about the Sun. The central theme of his masterful play *Copenhagen* is repeated here over and over in intellectual garb. "The ambiguity is the essence," he writes in a chapter entitled 'Mailing a cat'.

Uncertainty, ambiguity, complexity. How do we really know why there was a change in the weather? How did we decide whether to put honey or marmalade on our toast on Sunday morning? And why, next Sunday, might we make the opposite decision? How do I know that the brownish elliptical shape I see before me is a penny? Why do I think that I am me and you are you?

Frayn does not say the endeavour of science is empty. Rather, he says that we, the living observers, are an inseparable part of that endeavour. Furthermore, the Universe, including the grey matter between our ears, is far more complex than we can possibly understand.

Consider time, that most solid pillar of reality. According to Frayn, time does not exist without events, and events, in turn, make no sense without observers: "The probabilistic universe of quantum mechanics makes no sense unless it is inhabited."

Or consider language. Here, Frayn challenges Noam Chomsky's theory of a hard-wired universal grammar, writing: "Nothing in Chomsky's view accounts for the way in which all the rules of language are endlessly extended, adapted, and stood on their heads."

Quoting extensively from the physicist Niels Bohr, philosophers Daniel Dennett and John Searle, psychologist William James, philosopher of science Karl Popper, and physicist and philosopher of science David Deutsch, and indulging not a little in his own training in philosophy, Frayn tackles all these questions with erudition, wit, a spirited and inviting voice, and supreme self-confidence. Showing no fear, he takes on the leading thinkers in any number of fields. One cannot help but be impressed by the sheer ambition of the endeavour.

In large part, *The Human Touch* is about what its title suggests. But the ubiquitous references to physics make it clear that Frayn is talking about the external Universe as well. And here, I believe, he runs into some trouble and misunderstandings.

For example, Frayn says the uncertainty principle derives from first principles and "could have been hit upon before it was necessary to explain anything, simply from an abstract analysis of the concepts of position and velocity". (Frayn's own explanation of the principle in a footnote is erroneous, by the way.) This misunderstanding is not a small point because

of Frayn's repeated suggestion that the order of the physical Universe is largely a mental construction. The uncertainty principle never could have been reached by thought alone because it requires the concept of the quantum, which is absent from the classical understanding of position and velocity. The quantum was thrust upon the human mind at the beginning of the twentieth century by the need to explain experimental results, namely the observed black-body radiation from hot, dark containers.

Likewise, Frayn says that Einstein's theory of gravity would have remained unassailable even if its predictions had not been fulfilled, because it too was derived from first principles. The only conclusion of a negative experimental result, Frayn says, would have been that those principles were deemed not applicable to those particular situations. However, the equivalence principle, on which Einstein's general theory of relativity is based, could have been proven wrong by experiment, just as the concept of parity conservation was proven wrong in the 1950s and subsequently abandoned. First principles in science, no matter how beautiful and alluring, are eventually discarded if they repeatedly disagree with experimental results.

If Frayn is right that the logic and order of the external Universe are only products of the human mind, how can we separate the physical theories we conceive, such as quantum chromodynamics or the general theory of relativity, from ourselves, the conceivers? Don't the laws of nature, Frayn argues, simply reflect our human way of understanding? I don't think so. Suppose an intelligent civilization emerges somewhere in the Universe, produces scientists, conceives laws of nature describing such things as electricity and relativity, and then self-destructs. Before annihilation, the forward-thinking beings bury their physics books. Eons pass without any life forms in the universe. Then, ten billion years later, a new civilization emerges and develops its own equations for the laws of nature. One day, they accidentally uncover the physics books left behind by the previous society. They spend years deciphering the unfamiliar language but finally succeed, as modern humans did with the Rosetta Stone, and compare their theories of nature. What is the result? Every scientist I know would bet a roomful of pocket calculators that the separate descriptions of nature, composed billions of years apart, would have substantial agreement. The order and logic of nature, by whatever name you call it, seems to exist outside our heads.

If Frayn truly believes that the Universe has no logic beyond what we give it, then why does he repeatedly board aeroplanes for trans-Atlantic trips? Each time he flies, whether consciously or not, he is trusting his life to the logic and repeatability of nature, to the predicted relationships between air pressure, density, velocity, wing area and gravity.

I find Frayn to be at his most convincing

when describing the ambiguities and complexities of the human mind and heart. Here, his experience as a novelist and playwright, his insights into human psychology, and his good sense serve him well.

"Like me, I know, you are an endlessly flowing and shifting river of perceptions and feelings, and ever-springing source of decisions and initiatives. How can I ever lay hold of this subjectiveness?... We are like fetishists, who simplify the objects of their desire by reducing the intolerably complex whole to a pair of shoes or a piece of underwear."

In the end, I was moved by this book not

because of its science, or even its philosophy, but because of its poetry. *The Human Touch* is an extended ode to our humanity, to what it means to be a thinking being in this strange cosmos we find ourselves in. I was taken on a journey by a thoughtful companion. I was entertained. I was provoked. And I was left asking afresh some ancient questions. What is thinking? What is feeling? What is that intense sense of self that we have? What does it all mean? ■

Alan Lightman, a physicist as well as a novelist, is adjunct professor of humanities at the Massachusetts Institute of Technology and the author of *Einstein's Dreams* and *The Diagnosis*.

The beginning of wisdom

Is Pluto a Planet? A Historical Journey Through the Solar System

by David A. Weintraub

Princeton University Press: 2006. 248 pp.
\$27.95, £17.95

Stuart Ross Taylor

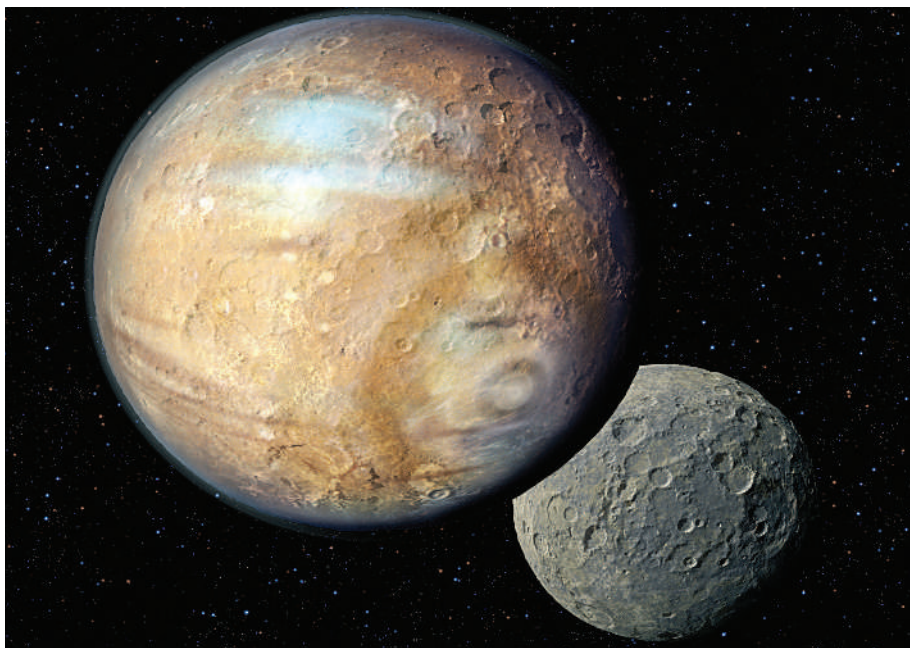
Pluto carries much the same sentimental, emotional and historical overload as Father Christmas. Even the name for this frozen dwarf, suggested in 1930 by 11-year-old Venetia Burney in Oxford, UK, is evocative, having mythological connections. And one of man's best friends appears as Disney's Pluto of the same vintage. It all makes for a heady cocktail.

Discovered by accident by Clyde Tombaugh, Pluto is a legacy of the obsession of Percival Lowell (he of the martian canals). It might otherwise not have been discovered for another 50 or 60 years, in which case the question of whether Pluto is a planet might never have arisen.

The discovery of Pluto's relatives, the trans-

Neptunian objects (TNOs), or Kuiper Belt objects as they are also known, is reminiscent of the history of the asteroid belt. Ceres, located in 1801, was hailed as the missing planet between Mars and Jupiter that was long predicted by the Titius-Bode law. But its trivial size and the rapid discovery of Pallas, Juno and Vesta in the vicinity suggested that the true situation might be more complex. It was not until 1845 that other bodies were discovered, and better telescopes led to a deluge, relegating the four new planets to the status of minor planets or asteroids. As they did not display a common origin point, even the notion that they were remnants of a shattered planet vanished.

Why, then, did it take so long to discover the predicted swarm of TNOs beyond Neptune? Probably because Pluto has such an eccentric and inclined orbit that no one was dedicated enough, or had sufficient funding or energy, to repeat Tombaugh's arduous search technique. Although Pluto's moon, Charon, was found in



After 76 years, astronomers still argue over the status of Pluto (represented here with its moon Charon).

D. VAN RAVENSWAAY/SPL

1978, it took David Jewitt and Jane Luu until 1992, searching with modern instruments, to discover the next body in the Kuiper Belt — a classic case of finding a needle in a haystack. Now we are overwhelmed with TNOs.

The story is well told in *Is Pluto a Planet?* by David Weintraub, which provides a readable historical account of our knowledge of the Solar System and the concept of what has been considered to be a planet. Fashions have changed with time, and the number of planets has ranged from 6 to 24. Specialists will note that the author passes over the questionable acquisition by Johannes Kepler of Tycho Brahe's data, which Kepler himself admitted to "usurping". And Uranus was discovered after a careful search of the heavens by William Herschel, arguably not by accident, but it was probably inevitable. I also found no mention of Burney, who now forms part of the mythology.

A concluding appendix gives a useful discussion of the properties of Pluto and Charon, preceded by the views of various astronomers and planetary specialists on what should constitute a planet. As always, these reveal as much about the commentator as the problem. I found the most perceptive was from Brian Marsden: "It has rarely been scientifically useful to use the word [planet] without some qualification."

Towards the end of this interesting book, Weintraub surprisingly concludes, despite the close analogy between the discovery of the asteroid and Kuiper belts, that we should retain Pluto as a planet by using the three physical parameters of orbital characteristics, mass and roundness. Although he admits that "we know that Pluto earned its status as a planet by accident", this suggestion accords Pluto a kind of royal status. The application of his criteria leads him to a total of 24 planets. This number is bound to grow and the classification is too broad to be scientifically or even culturally useful. As *The New York Times* has remarked, "too many planets numbs the mind".

Trying to define a planet runs into the philosophical difficulty of attempting to classify any set of randomly assembled items. A bewildering array of objects formed in the early solar nebula, including dust, asteroids, Trojans, centaurs, comets, TNOs and our eight planets, ranging from tiny Mercury to mighty Jupiter. All differ from one another in some salient manner, as do the 160 satellites. The key question is how did they form and evolve, not what pigeonhole can they be forced into. So the decision of the International Astronomical Union that there are eight major planets and a host of minor planets seems an appropriate compromise. But even so, qualifying terms, such as ice giants and Earth-like planets, will always be needed. We need, then, to recall the wise words of Confucius: "The beginning of wisdom is to call things by their right names."

Stuart Ross Taylor is in the Department of Earth and Marine Sciences, Australian National University. He is the author of *Solar System Evolution*.

The next pandemic



R. STEVENSON/AP/EMPICS

Is this your Christmas dinner? Farmed flocks of poultry could become infected with H5N1 flu.

Bird Flu: A Virus of Our Own Hatching

by Michael Greger

Lantern Books: 2006. 416 pp. \$30, £21.99

John Oxford

I am often kicked around by American authors in books about influenza. How dare a Limey suggest that the Spanish influenza A H1N1 virus arose in a gas-infected, pig-ridden and bird-infected army camp of 100,000 people in France in 1916, when the whole world knows it started in Dorothy's home state, Kansas? But I felt less bruised than usual. Perhaps I am getting used to it.

Bird Flu by Michael Greger pulls and pushes the reader along, and every virologist under the Sun is thrown into the mix. We are projected back to pandemics in 1918, 1957 and 1968. But mostly the searchlight illuminates the current and worrying spread of H5N1, the recent theories of pathogenesis, mutation and spread. Greger covers the science rather well, and his descriptions of the polymerase chain reaction, innate immunity, cytokine storms and virus evolution are as clear as crystal. He also provides hundreds of references to burrow into.

But does the book really deliver on its title? It certainly goes to town on the subtitle, burning the bridges that link us to the great international avian reservoir. With the exception of virologists Graeme Laver and Robert Webster, we have only recently realized that influenza A

is an avian virus securely seated in the 50 billion ducks, geese and swans that migrate across our planet. Periodically the virus makes the leap to infect humans. But not many of us stroke the feathers of a banded goose, so the virus first moves laterally to domesticated birds and then to their keepers.

The bridge leads to the industrialized chicken industry. I sympathize here. Unbeknown to the UK Department for Environment, Food and Rural Affairs, my wife keeps seven precious chickens at the bottom of the garden. Every autumn they are covered up to avoid the virus-spewing migratory geese who pass overhead. I have slowly cottoned on to why we pour grain into their hungry mouths. We don't like seeing chickens used as fodder, being scalded to death, chopped up alive or, worse, dying of H5N1. There can be no more unpleasant death than a neurological disease that kills a bird in 72 hours. The book is a hybrid, and the trick is to balance a scientific exposition about influenza with a cry from the heart about industrialized farming and how it demeans us. The author notes that "it may take a pandemic with a virus like H5N1 before the world realises the true cost of cheap chicken".

Greger is a clever writer. The book is a zinger, not deep like John Barry's *The Great Influenza* (Viking, 2004; see *Nature* 429, 345; 2004), but more worldly, broader and more scientific. It is also more global in its approach than most US books, but I suppose it has to be as the United

States has not yet experienced H5N1.

Fortunately, the world has woken up to the threat of H5N1. The US government has thrown \$9 billion at the problem, much more than against smallpox and polio combined. Every related research programme in the United States will benefit. The ripples have even reached Britain. There is now an axis of flu research, but will we join it? Yes please!

However, the book fails to confront the question I am asked daily: "Why are you so worried about 151 deaths from H5N1?" Well, go back to 1916, to Etaples in northern France, where a form of flu causing heliotrope cyanosis (a characteristic lavender coloration of the face) with

a case fatality of 60% was beginning to spread. There were 145 cases. At some point in the next two years it mutated to become more infectious and 30 times less virulent. Then it killed 50 million people. Doesn't this ring a nasty bell?

Greger focuses on the chicken industry. His descriptions of how awful it is make for tearful reading. I really mean this. He would do well to fire a few extra missiles at the political class. In 1916 their political antecedents were also preoccupied with war and other vainglorious affairs. They were too busy to read the warnings in *The Lancet* that a new disease was lurking. Even in Edwardian times, governments could have prepared themselves. The scientists

of the day reacted quickly enough; they made pneumonia vaccines, used masks, hygiene and social distancing, and many died in the attempt. Today we are well served in Britain and the rest of the European Union, but many nations in the world deserve better. They will face the next pandemic truly naked and bereft.

I am not as pessimistic as Greger. I like Churchill's attitude: "Give us the tools and we will finish the job." Well, for the first time in our history, we have them, and we will. ■

John Oxford is at the Centre for Infectious Diseases, Retroscreen Virology, Queen Mary, University of London, 327 Mile End Road, London E1 4NS, UK.

The snowflake man

Wilson Bentley's photographs of snow crystals strike a chord with us all.



Martin Kemp

Since its debut in Johannes Kepler's book *On the Six-Cornered Snowflake* in 1611 (see *Nature* **432**, 953; 2004), the snowflake shape has become as familiar as nature's other geometrical designs, such as the spider's web, the honeycomb and mollusc shells. The intricate and remarkably varied structures of snow crystals were shown for the first time in Robert Hooke's *Micrographia* in 1665.

The most unlikely hero in the later story of snowflake structure is a Vermont farmer with no orthodox scientific credentials, Wilson Bentley (1865–1931). Born into a farming family, the young Bentley developed an intense fascination with the structure of snow. At the age of 20, using a specially constructed rig of bellows camera and microscope, he took the first photograph of a single snow crystal.

One of the great eccentrics of science, 'Snowflake' Bentley (as he was known in his local community) resided steadfastly in the family farmhouse for his whole life. But this did not stop him publishing a series of well-regarded books and articles in scientific and popular journals.

His 1931 book *Snow Crystals*, which contains photographs (like that shown on the left) of almost 2,500 crystals, stands alongside Ernst Haeckel's illustrated volumes on the radiolaria as a classic on the geometrical sciences of nature. Bentley's systematic procedures, delicate manual dexterity, incredible patience and tolerance of working in freezing conditions were motivated by a sense of poetic wonder.

Some of his enraptured tone is conveyed in a piece he wrote for *Harper's Monthly Magazine*: "Quick, the first flakes are coming; the couriers of the coming snow storm. Open the skylight, and directly under it place the carefully prepared blackboard, on whose ebony surface the most minute form of frozen beauty may be welcome from cloud-land. The mysteries of the upper air are about to reveal themselves, if our hands are deft and our eyes quick enough."

It was Bentley who told the world that no two snowflakes are precisely alike. This variety, we now know, has a number of causes: the construction of the crystals from about 10^{18} water molecules; their formation under varying temperatures as they swirl with their cloud; and the unpredictability of the processes of aggregation from one crystal to another.

Bentley also knew that, in contrast to popular belief, most crystals are not wholly symmetrical — it's just that the symmetrical ones are generally selected for illustration.

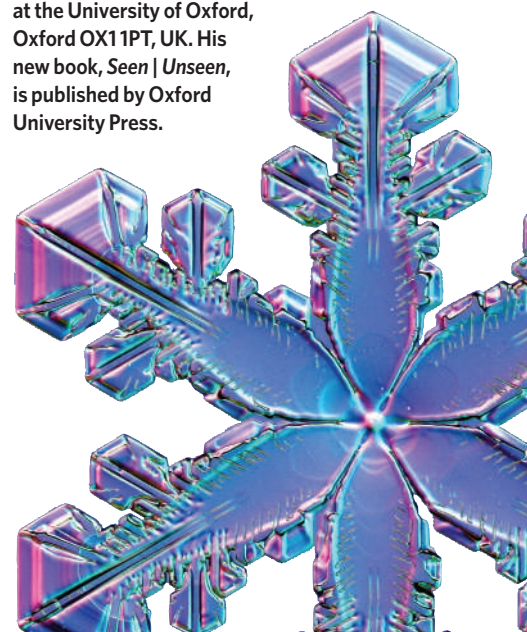
The complexity of the infinite variations on the hexagonal structure continues to attract serious scientific attention, for example from physicist Kenneth Libbrecht at the California Institute of Technology. Libbrecht's lively website, which includes animated images of crystal growth (www.its.caltech.edu/~atomic/snowcrystals/movies/movies.htm), is both informative and appropriately entertaining for the festive season. The

image shown below is from Ken Libbrecht's *Field Guide to Snowflakes* (Voyageur, 2006).

Mathematically, the snowflake has also been enlisted in the cause of fractals. In 1904 the Swedish mathematician Niels Fabian Helge von Koch published his 'snowflake curve', generated from an equilateral triangle. Each side is trisected and the centre segment replaced by two sides of a smaller equilateral triangle projecting outward. This process is repeated *ad infinitum*. The total length increases with each step, but the length of each side approaches zero. In this sense it is unlike a snowflake in nature, the lower limit of which is set by the size of a water molecule.

The hexagonal configurations familiar from Christmas lights in streets around the world and glittering on countless Christmas trees testify that the passion shown by these snowflake pioneers is shared by everyone who responds intuitively to the geometry of nature. Perhaps that is all of us.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK. His new book, *Seen | Unseen*, is published by Oxford University Press.



BUFFALO MUSEUM OF SCIENCE

K. LIBBRECHT/VOYAGEUR PRESS

NEWS & VIEWS

PHYSIOLOGY

Obesity and gut flora

Matej Bajzer and Randy J. Seeley

The intestinal bacteria in obese humans and mice differ from those in lean individuals. Are these bacteria involved in how we regulate body weight, and are they a factor in the obesity epidemic?

Much has been said and written about the sequencing of the human genome, and the research avenues that it has opened up. But our own genome is not the only one with which we need to be concerned. In particular, trillions of bacteria — collectively referred to as the microbiota — reside in our gastrointestinal tracts (Fig. 1), and each brings its own genome to this genetic party.

Reports by Gordon and colleagues^{1,2} on pages 1022 and 1027 of this issue explore the potential relationship between the types of microbiota found in the gut and the regulation of body weight. Although there is no doubt that human genetics plays a large part in determining body weight³, it is equally undisputed that the increase in prevalence of obesity over the past 25 years cannot be attributed to changes in the human genome⁴. The inference is that other factors are responsible, such as the availability of inexpensive, calorically dense foods, or the reduction in physical activity in our daily lives. The work described by Gordon and colleagues^{1,2} raises the possibility that our gut bacteria are another factor that contributes to differences in body weight among individuals.

Diverse evidence points to the importance of biological control systems that result in a close match between caloric intake and caloric expenditure. For the vast majority of humans (including obese individuals), caloric intake exceeds caloric expenditure by less than 1%, but even these small differences can accumulate over years to lead to detrimental increases in body weight⁴. The body's ability to match caloric intake to caloric expenditure is the result of the brain's ability to monitor the amount of fat in the body through changes in the levels of circulating hormones. One such hormone is leptin, the levels of which increase with increasing body fat⁵. Leptin deficiency in mice and humans results in unrestrained caloric intake and low caloric expenditure, with a consequent rapid increase in body weight⁶. Similarly, falling leptin levels are a primary reason for the difficulty in maintaining weight loss.

The two reports^{1,2} compare the genetic material collected from the microbiota in the gut of lean and obese mice and humans to

assess the relative abundance of various types of bacterium. The two predominant populations of microbiota in both the mouse and the human gut are members of the bacterial groups known as the Firmicutes and the Bacteroidetes (Box 1, overleaf). The authors¹ studied a small number of obese humans, and found that the proportion of the genetic material from Firmicutes was higher than the proportion in lean individuals. Moreover, when obese individuals lost weight over a year, the proportion from Firmicutes became more like that of lean individuals.

In the second report², the authors describe how differences in the microbiota of lean and obese mice confirmed this result, with obese mice having a higher proportion of intestinal Firmicutes. Importantly, the microbiota of obese mice was rich in genes encoding enzymes that break down otherwise indigestible

dietary polysaccharides. These differences seem to have functional consequences, as obese mice had more fermentation end-products and fewer calories remaining in their faeces than lean mice. Thus, the bacteria in obese mice seemed to assist their host in extracting extra calories from ingested food that could then be used as energy.

These two results collectively suggest that obesity alters the nature of the intestinal microbiota, but they do not prove that different relative proportions of bacteria can lead to different body weights. To test this possibility, the authors performed a clever experiment in which the microbiota of obese, leptin-deficient mice was transferred to lean, microbe-free recipient mice². Over a two-week period, mice given the microbiota from obese mice extracted more calories from their food and had a modest fat gain that was statistically



Figure 1 | Home for an abundant microbiological flora. The human gut and (inset) a scanning electron micrograph of part of the small intestine, with some bacterial inhabitants picked out in green.

BIOMED. IMAGING UNIT, SOUTHAMPTON GEN. HOSP./SPL

3DCLINIC/GETTY

Box 1 | The Firmicutes and the Bacteroidetes

The Firmicutes and the Bacteroidetes are divisions — or phyla — within the domain Bacteria. The microbiota of the human gut is dominated by their members, most of which are benign, although a few are pathogenic.

The Firmicutes is the largest bacterial phylum. It contains more than 250 genera, including *Lactobacillus*, *Mycoplasma*, *Bacillus* and *Clostridium*. There is considerable variety in the phylum. For example,

the *Clostridium* species are obligate anaerobes (that is, they absolutely require anoxic conditions), whereas members of *Bacillus* form spores and many of them are obligate aerobes.

Streptococcus pyogenes, the well-known cause of 'Strep. throat', is also a member of the Firmicutes.

The Bacteroidetes include about 20 genera. In the human gut, *Bacteroides* is probably the most abundant single genus, and *Bacteroides thetaiotaomicron*

is one of the most abundant organisms. Species of *Bacteroides* are obligate anaerobes that are benign inhabitants of the gut. However, they are opportunistic pathogens that can cause disease if they gain access to the peritoneal cavity, for example following surgery or a perforated ulcer. Members of the Bacteroidetes are found in the intestinal tracts of many warm-blooded animals, but are also abundant in soil and sea water.

M.B. & R.J.S.

greater than that of mice receiving microbiota from lean mice. Taken together, these data suggest that differences in the efficiency of caloric extraction from food may be determined by the composition of the microbiota, which, in turn, may contribute to differential body weights.

This is a potentially revolutionary idea that could change our views of what causes obesity and how we depend on the bacteria that inhabit our gut. But a great deal remains poorly understood. Most notably, it is not clear whether such small changes in caloric extraction can actually contribute to meaningful differences in body weight. There are few data that substantiate the predicted increased caloric extraction in obese humans. Small but persistent increases in efficiency might potentially cause the accumulation of excess body weight over long periods, but these small differences are not the cause of obesity in leptin-deficient mice. These mice have a single gene mutation that prevents the production of biologically active leptin⁷. The resulting increased caloric intake and reduced caloric expenditure is many times larger than the small difference in extraction that could be produced by differences in the microbiota. In fact, the differences in body fat between mice given the 'obese microbiota' and those given the 'lean microbiota' are so small that they could be accounted for entirely by the tiny differences in food intake, rather than by differences in caloric extraction.

Another unknown is why and how the make-up of the microbiota is shifted by differences in body weight. Given that acquiring food from the environment can be both calorically expensive and potentially dangerous, it would seem to be most adaptive to extract as many calories from every bite of food as possible. Moreover, if caloric extraction does become more efficient, the regulatory system would dictate that the organism responds by reducing its caloric intake. If a host organism had the ability to change its microbiota so as to increase caloric extraction, it would seem most adaptive to do so when facing famine conditions and losing weight. However, the data

indicate just the opposite — the microbiota seems to be more efficient in obese humans who already have the most stored energy, and shifts to being less efficient as the subjects lose weight¹.

There is also the issue of how conditions in the host organism could change the make-up of the microbiota. Low levels of leptin are a signal of starvation that triggers several changes in the neuroendocrine system that work to conserve calories⁶. Consequently, it would make sense that low leptin might also impart a signal to the microbiota to become more efficient at extracting calories from food. This hypothesis would fit the microbiota of the obese, leptin-deficient mice. However, in humans where

obesity is associated with increased leptin, the same trends in microbiota composition are found, making it unlikely that the microbiota is responding to leptin directly. Moreover, it seems that when the bacteria are transferred to a lean mouse in which leptin is normal, the bacteria retain their 'obese' character over a two-week period. Thus, it is not clear how gut bacteria 'know' whether the host is obese or lean.

Gordon and colleagues' results^{1,2} tempt consideration of how we might manipulate the microbiotic environment to treat or prevent obesity. But questions about how and why the composition of gut microbiota is regulated will have to be answered first. As we have discussed, those questions are many and various. The two papers nonetheless open up an intriguing line of scientific enquiry that will ally microbiologists with nutritionists, physiologists and neuroscientists in the fight against obesity. ■

Matej Bajzer and Randy J. Seeley are in the

Department of Psychiatry, Genome Research

Institute, University of Cincinnati, Cincinnati,

Ohio 45237, USA.

e-mail: seeleyrj@ucmail.uc.edu

1. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. *Nature* **444**, 1022–1023 (2006).

2. Turnbaugh, P. J. *et al. Nature* **444**, 1027–1031 (2006).

3. Farooqi, I. S. & O'Rahilly, S. *Endocr. Rev.* doi:10.1210/er.2006-0040 (2006).

4. Hill, J. O. *Endocr. Rev.* doi:10.1210/er.2006-0032 (2006).

5. Ahima, R. S. & Flier, J. S. *Nature* **382**, 250–252 (1996).

6. Woods, S. C., Benoit, S. C., Clegg, D. J. & Seeley, R. J. *Science* **280**, 1378–1383 (1998).

7. Zhang, Y. *et al. Nature* **372**, 425–432 (1994).

ASTROPHYSICS

A burst of new ideas

Bing Zhang

Gigantic cosmological γ -ray bursts have fallen into a dichotomy of long and short bursts, each with a very different origin. The discovery of an oddball burst calls for a rethink of that classification.

The events known as γ -ray bursts (GRBs) are the most violent and luminous explosions observed in the Universe. In the early 1990s, it became clear that they come in two distinct flavours: longer-duration bursts, typically longer than 2 seconds, with a spectrum of emitted radiation that peaks at lower ('softer') energy; and shorter-duration bursts with a more energetic, 'harder' spectrum¹. Observations of burst afterglows in the past decade — particularly in the past year^{2–4} — have seemed to show that this division is a clean one, and is firmly rooted in the progenitor of each type of burst. According to this picture, long bursts are associated with a young stellar population, marking the deaths of massive stars whose lifetime is short⁵. Short bursts, on the other hand, are associated with an old stellar population, and are

probably powered by mergers of compact objects such as neutron stars or black holes⁶.

In this issue, four papers^{7–10} blow a hole in this cosy paradigm. They contain observations of a bright γ -ray burst, GRB 060614, that triggered NASA's GRB sentinel, the Swift satellite, at 12:43:48 UT on 14 June 2006. The burst defies pigeonholing within the current scheme.

Gehrels *et al.* (page 1044)⁷ detail the circumstances of this peculiar burst's discovery. It is one of the brightest ever seen, and was soon located precisely not only by Swift's instrumentation, but also by other space- and ground-based telescopes. The burst is situated in the suburbs of a faint and relatively nearby dwarf galaxy⁸. Its duration, recorded by Swift as 102 seconds⁷, characterizes it unambiguously

as a long GRB. According to previous experience, evidence for the death of a star in a stellar explosion — a supernova — should have been spotted in the burst's neighbourhood before too long. But the many optical telescopes around the world trained on the target, waiting for yet another confirmation of the connection between GRBs and supernovae, saw nothing.

Three further papers^{8–10} provide independent reports of the stringent upper limits on the radiation flux from a possible supernova underlying GRB 060614. Gal-Yam and colleagues (page 1053)⁸ made a series of observations with the Hubble Space Telescope in the weeks after the burst trigger. These set an upper limit more than 100 times fainter than the faintest supernova previously associated with a GRB — and indeed considerably fainter than any supernova ever observed. Della Valle *et al.* (page 1050)⁹ report complementary observations from the European Southern Observatory's Very Large Telescope in the Atacama desert in northern Chile. This survey started 15 hours and ended 65 days after the burst, and provides an upper limit on the flux that is about three times higher than that of Gal-Yam and colleagues', but still well below the luminosity of any known supernova over an unprecedentedly long time span. Fynbo *et al.* (page 1047)¹⁰ use a range of telescopes to arrive at a similar result — and also discover a second long burst with no apparent supernova signature.

The absence of a supernova need not in itself be revolutionary. The production of a significant amount of nickel-56, which is a prerequisite for a supernova, is not guaranteed in a collapsing star^{7,8}, and the earliest model to connect GRBs with the massive-star collapses indeed characterized the bursts as 'failed supernovae'⁵. A supernova might also precede its associated GRB¹¹. Nevertheless, the weight of evidence from the past decade is consistent with there being no significant gap between a GRB and its supernova, as well as with the hypothesis that every long GRB has a supernova accompanying it¹².

What makes the story more intriguing is that every property of GRB 060614 places it in the short-burst category — except, that is, for its duration. Gehrels *et al.*⁷ show that the time-lag of softer radiation behind harder parts of the burst's spectrum is small, as it is in a short GRB. Gal-Yam *et al.*⁸ find that the afterglow of the burst has a large offset from the star-forming region of the host galaxy; again, a feature more characteristic of a short GRB. Similarly, the star-forming rate of the host galaxy is relatively small compared with those of normal long GRBs^{8–10}, consistent with an old stellar population more likely to host a short burst.

Even regarding its duration, the seemingly long GRB 060614 can be shortened. Its γ -ray light curve consists of a short, hard early episode lasting around 5 seconds, followed by a long, soft tail⁷. Recent observations also indicate that most 'short' GRBs are not necessarily

	Type Ia	Type II (Type Ib/c)
Stellar population	Old	Young
Host galaxy	All types of galaxy	Late-type galaxies
Progenitor	Binary systems (accretion-induced collapse of white dwarfs)	Single-star systems (core collapse of massive stars)

Figure 1 | The classification of supernovae¹⁸.

	Type I (short-hard)	Type II (long-soft)
Duration	Usually short (may have a long tail?)	Usually long
Spectrum	Usually hard (tail is soft)	Usually soft
Spectral lag	Short	Long
Associated supernova	No	Yes
Stellar population	Old	Young
Host galaxy	All types of galaxies (predominantly in regions of low star formation rate)	Late-type galaxies (predominantly in irregular, dwarf galaxies)
Location in the host galaxy	Outskirts	Central
Progenitor	Mergers of compact objects in binary systems?	Single-star systems? (Core collapse of massive stars)

Figure 2 | A classification scheme for γ -ray bursts. The red rows show the analogy with the supernova classification scheme. Blue cells show the properties of GRB 060614 (refs 7–10).

so short, and are usually followed by a softer emission tail lasting around 100 seconds^{3,4}. Using an empirical relation between the spectrum hardness and the total energy budget of GRBs, it is possible to show¹³ that GRB 060614 would be marginally classified as short if only it were around eight times less energetic.

So how can this burst and the existing classification scheme be squared? There are in principle three possibilities⁸. First, GRB 060614 is indeed a long GRB associated with a collapsing star. If so, its progenitor must be very different from those of most other long GRBs because of the anomalous properties detailed above. Second, the burst belongs to the merger-type short GRBs — in which case, these should not carry the name 'short' any more. Third, this is the prototype of a completely different, third category of burst.

Given that the long–short paradigm is no longer adequate to describe the entire GRB phenomenon, a new terminology can be invented. Dividing bursts into Type I and Type II bursts by analogy with the supernova classification scheme might seem to lack imagination, but a comparison shows that such a definition might not be a bad choice (Figs 1, 2)¹³. The progenitors of Type Ia supernovae, like those of the traditional short GRBs, belong to the old stellar population and live in binary systems. Similarly, the progenitors of type II supernovae, like those of traditional long GRBs, are collapsing massive stars that die at a young age.

Comparing the observational properties of GRB 060614 (blue cells)^{7–10} with the multiple criteria in Fig. 2, one can see that this burst merges the properties of both categories. If one prefers to fit this burst into the straitjacket of bimodal classification, as I do, it would seem safer to apportion it to Type I, the traditional

short category. There are great theoretical difficulties in producing extended radiation emission from a merger of compact stars, although various ideas to overcome these problems have been suggested^{14–17}.

To resolve definitively whether GRB 060614 is a peculiar example of a Type I burst, or whether it is a representative of a third class of object, more data are needed. In particular, given that our current information about the make-up of this GRB's host galaxy does not fully rule out a Type II origin, discovering whether or not GRBs with similar properties will be detectable in elliptical host galaxies^{8,13}, which consist entirely of older stellar populations, will hold the key to the final answer. ■

Bing Zhang is in the Department of Physics and Astronomy, University of Nevada, 4505 Maryland Parkway, Las Vegas, Nevada 89154-4002, USA.
e-mail: bzhang@physics.unlv.edu

- Kouveliotou, C. *et al.* *Astrophys. J.* **413**, L101–L104 (1993).
- Gehrels, N. *et al.* *Nature* **437**, 851–854 (2005).
- Fox, D. B. *et al.* *Nature* **437**, 845–851 (2005).
- Barthelmy, S. D. *et al.* *Nature* **438**, 994–996 (2005).
- Woosley, S. E. *Astrophys. J.* **405**, 273–277 (1993).
- Paczynski, B. *Astrophys. J.* **308**, L43–L46 (1986).
- Gehrels, N. *et al.* *Nature* **444**, 1044–1046 (2006).
- Gal-Yam, A. *et al.* *Nature* **444**, 1053–1055 (2006).
- Della Valle, M. *et al.* *Nature* **444**, 1050–1052 (2006).
- Fynbo, J. P. U. *et al.* *Nature* **444**, 1047–1049 (2006).
- Vietri, M. & Stella, L. *Astrophys. J.* **492**, L59–L62 (1998).
- Woosley, S. E. & Bloom, J. S. *Annu. Rev. Astron. Astrophys.* **44**, 507–556 (2006).
- Zhang, B. *et al.* *Astrophys. J. Lett.* (in the press); preprint available at www.arxiv.org/astro-ph/0612238
- Faber, J. A. *et al.* *Astrophys. J.* **641**, L93–L96 (2006).
- Dai, Z. G. *et al.* *Science* **311**, 1127–1129 (2006).
- Rosswog, S. preprint available at www.arxiv.org/astro-ph/0611440 (2006).
- King, A. *et al.* *Mon. Not. R. Astron. Soc.* (in the press); preprint available at www.arxiv.org/astro-ph/0610452
- Filippenko, A. V. *Annu. Rev. Astron. Astrophys.* **35**, 309–355 (1997).

NEUROBIOLOGY

Auditory fidelity

Thomas D. Parsons

Detailed investigation of a molecule involved in an inherited type of deafness reveals a fresh facet to the mammalian auditory system — a hitherto unknown way for synapses to put calcium in a bind.

Mammals react to sounds with exquisite temporal fidelity, a feat that is initiated by precise calcium-dependent signalling in the hair cells of the inner ear. Calcium is a common signalling molecule in the nervous system and elsewhere. Writing in *Cell*, Roux *et al.*¹ describe how they have identified an unexpected player in the auditory system. That player is a molecule called otoferlin, which has not previously been implicated as a calcium sensor for neurotransmitter release in nerve function.

A sensitive auditory system confers a tremendous evolutionary advantage, as it protects us from the things we fear most — those we cannot see. The ability to localize these potential dangers or communicate these possible threats depends on the precise timing of signals in the neural code that the brain ultimately perceives as sound. The basis of this temporal fidelity lies in the control of communication between the mechanosensory cell of the cochlea, the inner hair cell, and its downstream partner, the auditory nerve (Fig. 1). The mechanical energy of acoustic waves causes minute displacements of sensory hair bundles extending from the cell. These deflections result in rapidly oscillating electrical potentials, which trigger calcium influx from outside the cell; that in turn prompts tiny subcellular organelles filled with a chemical messenger to dump their contents into a well-defined extracellular compartment, the synaptic cleft. This messenger, or neurotransmitter, excites a nearby, afferent nerve fibre, which elicits an all-or-none electrical response that propagates details of the acoustic stimulus to the brain.

The capacity of the auditory system to follow acoustic waves oscillating at several thousand times per second suggests that this calcium-dependent regulatory event may need to be as much as ten times more precise than signalling between most types of neuron. So how does the inner hair cell achieve such exquisite temporal fidelity in its release of neurotransmitter? Roux *et al.*¹ suggest that the hair cell has evolved a unique calcium-sensing molecule, otoferlin, for controlling neurotransmitter release. The action of otoferlin allows a hair cell's specialized synapses — ribbon synapses, a specific class of afferent synapse common to sensory systems — to meet the requirements of hearing.

Both mice and humans suffer from an inherited form of deafness called DFNB9. Defects in otoferlin are responsible, and Roux and colleagues hypothesized that otoferlin might be

involved in the correct operation of the synapse between the hair cell and the afferent nerve fibre. They found that, in mice, not only is otoferlin localized to the synaptic vesicles of inner hair cells, but that it also undergoes developmental changes in expression concurrent with the formation of ribbon synapses. Otoferlin also binds in a calcium-dependent manner to SNARE proteins, highly conserved molecules thought to be essential for the release of neurotransmitters and for other events requiring fusion of membranes.

The authors genetically manipulated the otoferlin molecule to prevent its functional expression. Mice lacking both copies of the normal gene have structurally normal synapses between the hair cell and afferent fibre, but are deaf and lack calcium-triggered dumping of the synaptic-vesicle contents. Interestingly, only the most rapid phase of putative neurotransmitter release is abolished by a defect in the otoferlin molecule. This fast component is widely thought to be associated with a specialized class of the vesicles that are close to the

cell surface and molecularly poised for release (Fig. 1).

This paper¹ is of great interest to both specialists in hearing research and neuroscientists in general, for several reasons. First, the mystery of the molecular entity mediating the temporal fidelity of signalling by the primary synapse in the auditory system may now be solved. At most fast synapses, the transmembrane protein synaptotagmin I is thought to be the calcium sensor of fast, synchronized neurotransmitter release^{2,3}. Thus, the possibility that the hair-cell synapse, or perhaps even other synapses, use a different molecule for calcium sensing is intriguing.

Second, all synaptotagmin molecules have so-called C2 domains, putative calcium-binding regions that are proposed to be responsible for its calcium-sensing functions^{4,5}. Otoferlin contains six of these C2 domains, presumably for the binding of calcium, but it remains to be seen if or how other parts of the otoferlin molecule contribute to the unique properties of calcium-dependent signalling by the cochlear inner hair cell. Third, the new work raises the question of whether otoferlin or related molecules have a function at conventional synapses. Ferlin family members in non-neuronal cells have been identified as participants in membrane fusion events related to membrane repair⁶.

Finally, Roux and colleagues' experiments¹ show how difficult it is to ascribe specific functions to molecules essential to synaptic-vesicle cycling. It is puzzling that the inner hair cell

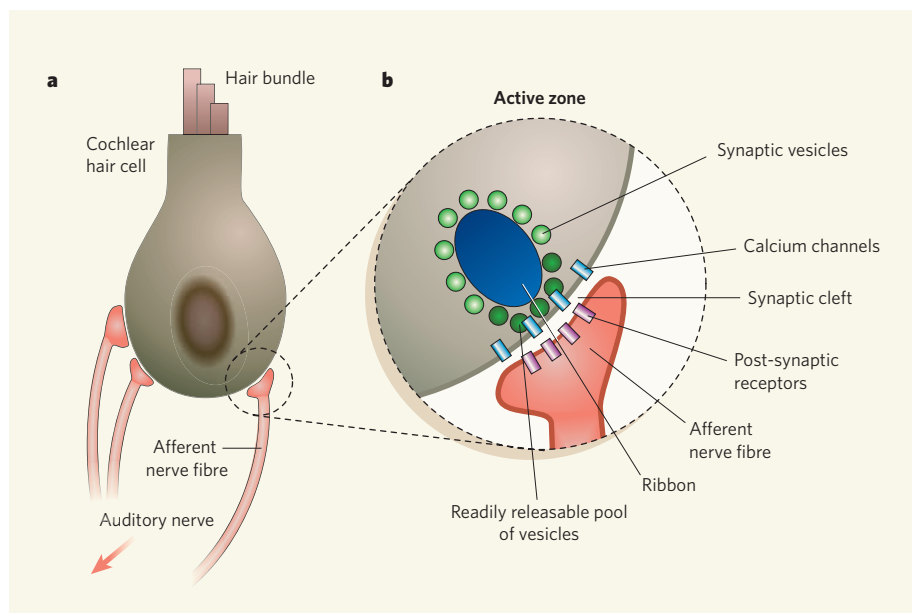


Figure 1 | Sound reception and perception. **a**, In the mammalian ear, the hair bundle on a cochlear hair cell senses variations in sound-induced pressure in the cochlea. The resulting, voltage-dependent signal is transduced and passed via the afferent nerve fibre and the auditory nerve to the brain, where it is perceived as sound. **b**, The voltage-dependent signal in the hair cell controls calcium channels, prompting calcium influx from outside the cell. Consequent changes in the cell's active zone result in neurotransmitter release into the synaptic cleft. In the active zone, aggregates of neurotransmitter-containing vesicles are tethered to a structure called the synaptic ribbon, localized to concentrated sites of calcium influx. A subpopulation of these vesicles lies close to the hair-cell membrane and constitutes a readily released source of neurotransmitter. Roux *et al.*¹ show that otoferlin is essential for the function of the auditory ribbon synapse, probably through its calcium-binding ability.

ribbon synapse looks so normal in mice with such a profound functional deficit. Given the blockade of synaptic-vesicle fusion, you would expect docked vesicles to accumulate at the release sites, but Roux *et al.* found no difference between the number of vesicles in normal mice and that in mice in which otoferlin production had been knocked out. The abolition of the fastest component of neurotransmitter release suggests that otoferlin acts on the most molecularly 'mature' vesicles, those conferred with rapid kinetics. Thus, otoferlin must mediate one of the final steps in the signalling cascade. However, definitive evidence of otoferlin's role as a calcium sensor

awaits mutagenesis experiments on its putative calcium-binding C2 domains.

Thomas D. Parsons is in the Auditory Research Laboratory, School of Medicine, and the School of Veterinary Medicine, University of Pennsylvania, 382 West Street Road, Kennett Square, Pennsylvania 19348, USA.
e-mail: thd@vet.upenn.edu

1. Roux, I. *et al.* *Cell* **127**, 277–289 (2006).
2. Chapman, E. R. *Nature Rev. Mol. Cell Biol.* **3**, 498–508 (2002).
3. Sudhof, T. C. *Annu. Rev. Neurosci.* **27**, 509–547 (2004).
4. Fernandez-Chacon, R. *et al.* *Nature* **410**, 41–49 (2001).
5. Rhee, J. S. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 18664–18669 (2005).
6. Doherty, K. R. & McNally, E. M. *Trends Mol. Med.* **9**, 327–330 (2003).

PARTICLE PHYSICS

Neutrons radiating decay

Nathal Severijns

That neutrons can be transmuted to protons, electrons and antineutrinos through the process of beta decay is old hat. That photons sometimes also get in on the act was suspected, but until now never confirmed.

The neutron, the existence of which was inferred by James Chadwick¹ in 1932, was one of the first members of nature's extensive particle zoo to be discovered. Yet it is only now, more than 70 years later, that one particular decay mode of the 'free' neutron — that is, a neutron not bound into an atomic nucleus — has been observed. On page 1059 of this issue, Nico *et al.*² describe how they caught up with this remarkably elusive phenomenon.

The 'radiative decay' that the authors observe is a consequence of the theory of quantum electrodynamics (QED), which describes all processes mediated by the electromagnetic force. Together with quantum chromodynamics,

which characterizes the 'strong nuclear' force, and the theory of the weak nuclear force, QED is one of the three pillars of particle physicists' 'standard model'. According to the theory, whenever one or more electrically charged particles are created in a particle decay, each one may from time to time emit a light particle, or photon, in a process known as inner bremsstrahlung. For a free neutron n — which is unstable, with a lifetime of some 900 s — this radiative process is written $n \rightarrow p + e^- + \bar{\nu}_e + \gamma$, where the decay products are (aside from a photon, γ) those present in the familiar weak-force-mediated beta decay of the neutron: a proton p , an electron e^- and a ghostly particle

known as an electron antineutrino, $\bar{\nu}_e$.

Although the radiative decay mode had previously been observed in the decay of radioactive nuclei and many elementary particles, it had never been observed in the decay of the free neutron. Only an upper limit on its probability existed³. Two recent theoretical works^{4,5} have predicted a branching ratio of about 3×10^{-3} for photon energies larger than 15 kiloelectronvolts (keV). This number indicates the fractional intensity of a particular decay mode compared with all possible decays; in other words, one would expect a photon to be produced in three out of every thousand neutron decays.

Nico *et al.*² observed the decay of neutrons — as is often the case in experiments with free neutrons — during their passage through the experimental apparatus. In order that the neutrons should spend sufficient time in the apparatus, 'cold' neutrons are typically used. These have energies in the range 0.1–1.5 millielectronvolts that correspond to velocities of 140–1,000 m s⁻¹. Neutrons this slow are only formed in the rare process in which thermal neutrons obtained from a nuclear reactor lose almost all of their energy in a single collision.

Nico and colleagues performed their experiments at the research reactor of the US National Institute of Standards and Technology in Gaithersburg, Maryland. Their set-up has three features that are crucial to observing the radiative decay of the neutron (Fig. 1). The first relates to the challenge of distinguishing the low rate of low-energy radiative-decay photons from the intense background rate of photons. This background rate is produced by the scattering and capturing of neutrons from the beam in the materials around the detectors. The authors sharply reduced the background count by covering the surfaces that could be hit by neutrons with materials containing lithium-6: only about 1 in 10,000 neutrons captured on lithium-6 give rise to a photon.

A second feature of the apparatus takes into account the fact that electrons produce photons when they are slowed down in a detector, in a process known as external bremsstrahlung. These photons are nearly indistinguishable from radiative-decay photons. Nico *et al.*² therefore counted electrons and photons with two separate detectors placed at opposite ends of a region permeated by a strong magnetic field. Because electrons, unlike photons, are charged particles, they spiral along these field lines, and so can be guided out of the neutron beam by a slight bend in the magnetic field towards one end of the apparatus. (The protons created in the neutron decay follow similar paths, and are counted in the same detector.) The distance between the electron and the photon detectors, as well as the shielding between them, significantly reduces not only the photon background in the electron detector, but also the number of external-bremsstrahlung photons created in the electron detector that can reach the photon detector.

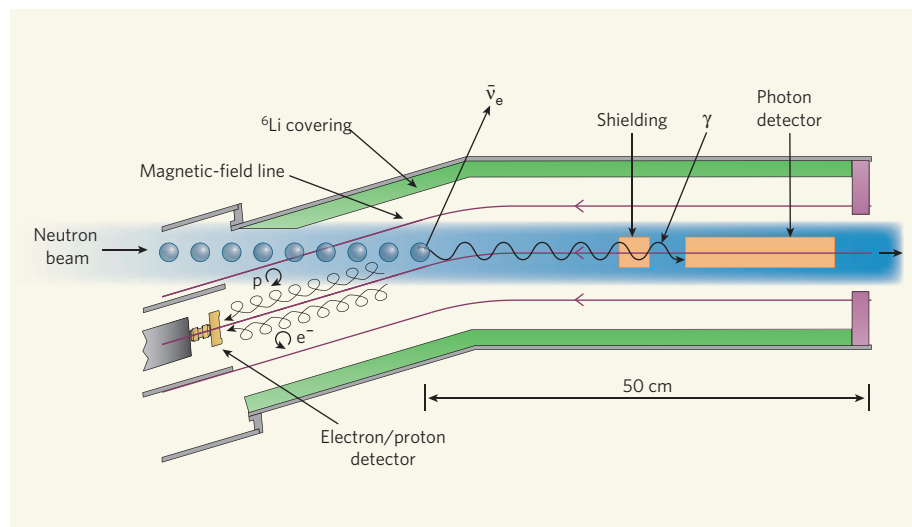


Figure 1 | Watching neutrons decay. Nico and colleagues' apparatus² included a number of features to ease the observation of the radiative decay of the neutron, including lithium-6-containing surfaces to minimize background photons, and separate detectors for protons and electrons, and for photons.

The third and final experimental detail arises from the characteristic signature of a radiative-decay event: the radiative photon arrives first, followed closely in time by the electron, with the proton, because of its greater mass, lagging far behind. Nico *et al.* could thus further reduce background events by requiring the two charged particles and the photon to be observed in their respective detectors in exactly this sequence, within a period of about 20 microseconds.

Having taken these and other potential disturbing factors into account, Nico and colleagues' analysis yielded² a branching ratio of $(3.13 \pm 0.11) \times 10^{-3}$ in the energy region between 15 keV and 340 keV that is entirely consistent with the theoretical calculations.

With the radiative-decay mode for the neutron thus established experimentally, new

opportunities arise. If the precision of such measurements is further improved — modifications of the apparatus to this end have already started² — an alternative means to check several crucial parameters of the standard model could be provided. These include the vector and axial-vector weak coupling constants, g_V and g_A , which determine the rate of neutron and nuclear beta decay³. The precision of quantities associated with free-neutron decay could also be improved to beyond the present level of one part in 1,000 at which radiative corrections become important. These corrections are currently calculated from theory, but the largest part of them (the 'outer' radiative correction) has now been confirmed by the result of Nico and colleagues².

Furthermore, a measurement of the circular polarization of the radiative-decay photons

could test the so-called Dirac structure of the weak current⁴ — that is, ascertain whether the neutrino and antineutrino really are two fully distinct particles. That in turn would tell us why interactions mediated by the weak force seem to maximally violate parity — nature's mirror symmetry — when this symmetry is respected by all other forces^{4,5}.

Nathal Severijns is at the Instituut voor Kern- en Stralingsfysica, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Leuven, Belgium. e-mail: nathal.severijns@fys.kuleuven.be

1. Chadwick, J. *Nature* **129**, 312 (1932).
2. Nico, J. S. *et al.* *Nature* **444**, 1059–1062 (2006).
3. Beck, M. *et al.* *JETP Lett.* **76**, 332–336 (2002).
4. Gaponov, Yu. & Khafizov, R. U. *Nucl. Instrum. Meth. A* **440**, 557–561 (2000).
5. Bernard, V., Gardner, S., Meissner, U.-G. & Zhang, C. *Phys. Lett. B* **593**, 105–114 (2004).

AGEING

Too fast by mistake

Tom Kirkwood

The intricate process of ageing involves numerous physiological pathways, together with genetic and environmental factors. Insight into this complex biology could come from studying a disorder that accelerates ageing.

Contrary to general expectation, human life expectancy in developed countries has not bumped into a ceiling, but continues to increase by around two years per decade¹ — or five hours per day. The reason previous forecasts proved wrong is simple. It used to be assumed that the ageing process was programmed, that some inner clock set a limit to life. But advances in the understanding of evolutionary mechanisms, and the experimental dissection of the genetics of ageing and longevity, reveal that this is not so². Instead, the genetic control of longevity comes through regulation of the body's survival mechanisms — those essential maintenance and repair processes that slow the build-up of the molecular and cellular faults that will eventually undo us (Box 1).

The indirect nature of this genetic control probably explains why the heritability of human longevity — the capacity for long life to run in families — is only modest³. More importantly, it explains the apparent malleability of the ageing process, revealed by the ongoing increase in life expectancy. If we can slow the onslaught of damage, or boost repair capacity, we can aspire to live longer, healthier lives without altering our genetic make-up. In this issue, Niedernhofer *et al.* (page 1038)⁴ report a new genetic disorder in humans that causes accelerated ageing, and the analysis of a mouse model of this condition, yielding valuable insights into ageing.

The condition was initially observed in a

15-year-old boy who had unusual sensitivity to sunburn and an array of clinical signs of accelerated ageing in many organ systems. The spectrum of symptoms differed from previously recognized skin-sensitivity disorders such as xeroderma pigmentosum, although there was some overlap. For example, ultraviolet light normally induces a DNA repair process, called nucleotide-excision repair,

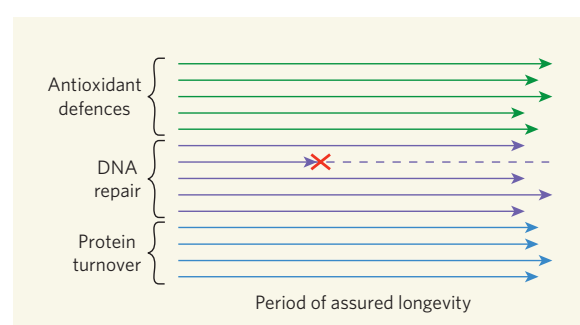
to deal with any damage caused by this radiation, but in both these ageing disorders this response is almost completely absent. It turned out that the mutation responsible for the new condition severely affected an enzyme called XPF-ERCC1 endonuclease. This endonuclease is necessary for the repair of damage that distorts the DNA helix or the crosslinks between DNA strands. The disorder adds to a growing list of inherited 'progeroid' syndromes — many linked to DNA-repair defects — that resemble, but do not exactly equate with, an accelerated process of normal ageing⁵.

So what, if anything, does this tell us about the biology of normal ageing? In the human disorder, the XPF subunit of the XPF-ERCC1 endonuclease is mutated. But to probe this question, Niedernhofer *et al.*⁴ exploited already available mice that lack the ERCC1 subunit (*Ercc1*^{-/-} mice). These animals show a

Box 1 | Multiple pathways affect longevity

Longevity is regulated by the joint actions of many maintenance and repair pathways, which control components of systems such as those involved in antioxidant defence, DNA repair and protein turnover. Each pathway guards against the age-related accumulation of particular kinds of damage and provides a period of 'assured longevity' before the relevant lesions build up to a level that threatens survival.

The plasticity of ageing and its susceptibility to environmental effects arises because many non-genetic factors, such as lifestyle and nutrition, can affect, for better or worse, the response mechanisms



that deal with exposure to damage. When a mutation reduces the effectiveness of one of these mechanisms, as shown in the figure by the red cross and truncated arrow, the corresponding lesions accumulate sooner, resulting in a shorter period of assured longevity, which may shorten

the lifespan of the organism as a whole.

Whether the resulting pathology resembles normal ageing will be governed by the extent to which these lesions contribute in a normal organism to the natural ageing process, albeit at a slower rate.

T.K.

pattern of accelerated degenerative changes very similar to those seen in the human patient, not only at the molecular and cellular level, but also in terms of organ and system functions. Comprehensive analysis of gene expression in the *Ercc1*^{-/-} mice compared with normal controls revealed a broad spectrum of changes, with 1,675 genes showing significantly different expression in the mutant mice. These changes included a general decrease in the activity of hormonal pathways involved in the regulation of metabolism, such as growth hormone/IGF1 signalling, and increased activity in anti-oxidant and DNA repair pathways.

Intriguingly, this pattern of altered gene expression in a model of accelerated ageing is reminiscent of the array of changes previously reported in the context of gene mutations that increase lifespan in the nematode worm *Caenorhabditis elegans* (a classic genetic model of ageing)⁶, and also during life-extending dietary restriction in mice⁷⁻⁹. Curiouser and curiouser, as Alice observed in Wonderland.

To resolve the paradox presented by these data, we need to grapple with one of the central conundrums in ageing research. When so many molecular and cellular changes occur in concert both during normal ageing and in genetic models of accelerated or postponed ageing, which of them are causative and which merely

secondary consequences? Niedernhofer *et al.* address this problem to some extent by showing that the profile of gene-expression changes in the *Ercc1*^{-/-} mice is broadly reproduced not only in naturally aged control mice, but also in adult control mice subjected to an increased burden of DNA damage from the chemical mitomycin C. From this, we can infer that the systemic changes in the *Ercc1*^{-/-} mice, and probably those in the XPF-deficient human condition, are a nonspecific consequence of the accumulated DNA damage that also occurs in ageing. But how particular are these changes to DNA damage, as opposed to the other kinds of damage that might affect ageing cells?

The collection of progeroid syndromes in mice and humans is intriguing, both for the similarity in age-accelerated symptoms (despite the diversity of causative mutations) and for the fact that these changes only partly mimic the normal ageing process¹⁰. And if, as various data indicate, reduced growth hormone/IGF1 signalling is a putative cause of increased longevity in mice, worms and fruitflies, how is it that such a pattern is also seen as a consequence of the damage that leads to the decreased longevity described by Niedernhofer *et al.*?

It is fast becoming clear that unravelling the complex biology of ageing so as to understand the true relationship between cause and effect

will not be easy. We will need to move beyond essentially descriptive studies of what goes up and what goes down, and perhaps to immerse ourselves in the exciting but deep waters of systems biology. This branch of science offers a way of teasing apart the complexity that allows a change in some components of a system to act as both cause and effect. Although the study of progeroid syndromes such as that due to the XPF-ERCC1 mutation poses at least as many questions as it answers, it also represents a helpful buoy in these waters. ■

Tom Kirkwood is in the Centre for Integrated Systems Biology of Ageing and Nutrition, Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne NE4 6BE, UK. e-mail: tom.kirkwood@ncl.ac.uk

1. Oeppen, J. & Vaupel, J. W. *Science* **296**, 1029–1031 (2002).
2. Kirkwood, T. B. L. *Cell* **120**, 437–447 (2005).
3. Finch, C. E. & Tanzi, R. E. *Science* **278**, 407–411 (1997).
4. Niedernhofer, L. J. *et al. Nature* **444**, 1038–1043 (2006).
5. Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H. & Vijg, J. *Science* **299**, 1355–1359 (2003).
6. Hansen, M., Hsu, A. L., Dillin, A. & Kenyon, C. *PLoS Genet.* **1**, 119–128 (2005).
7. Lee, C. K., Klopp, R. G., Weindruch, R. & Prolla, T. A. *Science* **285**, 1390–1393 (1999).
8. Cao, S. X., Dhabhi, J. M., Mote, P. L. & Spindler, S. R. *Proc. Natl Acad. Sci. USA* **98**, 10630–10635 (2001).
9. Spindler, S. R. *Mech. Ageing Dev.* **126**, 960–966 (2005).
10. Kipling, D., Davis, T., Ostler, E. L. & Faragher, R. G. *Science* **305**, 1426–1431 (2004).

OPTICS

A light touch

'Being good with your hands' is not quite the badge of approval it once was: the tiny scale of the components used in most high-tech industries has made old-fashioned manual dexterity largely redundant. Yet in many sensitive applications, the unique directness of human touch is highly desirable. Now Graeme Whyte *et al.* introduce an apparatus that could go some way to putting the 'hand' back in manipulation, even on very small scales (*Opt. Express* **14**, 12497–12502; 2006).

The authors' apparatus marries the skill of human digits with the laser-sharp precision of 'optical tweezers'. Such tweezers use the momentum of a highly focused laser beam to trap and move objects as small as a single atom. The technique is increasingly useful, especially in cell biology, where it can be applied to

measuring the mechanical, optical and transport properties of cells and biological molecules without the need for a material probe.

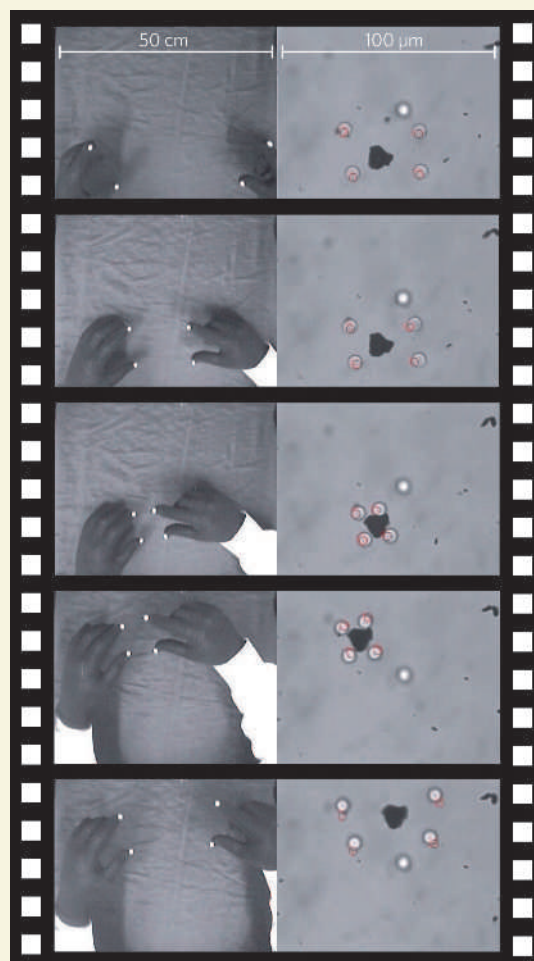
Whyte *et al.* use 'holographic' optical tweezers, in which a diffractive element, known as a spatial light modulator, is used to steer the laser beam. This allows several optical traps to be created that can be moved independently in three dimensions. The authors hook up the spatial modulator to a camera that images the position of beads attached to the fingertips of two gloved human hands. A specially written hologram-calculation algorithm converts the positions of the beads in the photographic plane into the x-y position of the optical traps, and the apparent size of the bead in the image into a z-coordinate.

The result is a device in which each trapping beam acts as a digit of an optically

controlled hand. The position of this hand on the micrometre scale is controlled by appropriate larger-scale movements of the remote, real hands. The images show how silica beads trapped at the tips of the optical fingers grasp and move a larger, irregularly shaped chrome bead. The hologram algorithm allows 8 frames per second — enough for manipulation in real time — to be processed using a standard desktop computer.

Because the manipulated object does not itself have to be optically trapped, this method could be very useful for dealing with light-scattering metallic particles or light-sensitive materials, such as some biological tissues. But more generally, such techniques could mean that the future of micromanipulation lies in our hands.

Richard Webb



OPT. EXPRESS

MATHEMATICS

Proof at a roll of the dice

Bernard Chazelle

The PCP theorem encapsulates the idea that randomization allows the immediate verification of any mathematical proof. A simple route to this striking result was proposed earlier this year.

Many branches of mathematics have their signature numbers: geometry has the transcendental π ; analysis Euler's exponential e ; and algebra the imaginary unit, the square-root of -1 , i . These numbers are the stars of exotic — but seemingly quite separate — worlds scattered across the vast expanse of the mathematical cosmos. Until, that is, one day a master stargazer spots a stellar alignment of breathtaking beauty, and nothing ever looks quite the same again. For these worlds, that alignment is Euler's identity, $e^{i\pi} = -1$, and it is proof, if proof were needed, that mathematics is about more than shape (π), change (e) and structure (i): it is also about magic.

Every science boasts its own brand of this wonder of unification: computer scientists, for their part, marvel at what they call PCP, shorthand for probabilistically checkable proof. Simply stated, this is the curious phenomenon that the mere ability to toss coins makes it possible to check the most complex of mathematical

proofs at no more than a passing glance. This remarkable fact was discovered by Arora *et al.*^{1,2} a decade ago, and proving it has relied on a vast arsenal of complex algorithmic and algebraic techniques. In the spring of 2006, however, Irit Dinur proposed an elementary proof³. In the short years of its existence, PCP has revolutionized the field of approximation algorithms — methods for finding nearly optimal solutions to problems that cannot be solved exactly within a reasonable time. Dinur's result is the latest chapter in one of the most engrossing chronicles of computer science.

To appreciate fully the significance of PCP, imagine you wake up one morning with your head full of a complete proof of the Riemann hypothesis. (This is arguably the greatest open problem in mathematics, and is a deep statement about the distribution of the prime numbers, the atoms of arithmetic.) Giddy with anticipation of certain fame, you leap out of bed and type up the proof, in its full 500-page

glory. That done, you are seized by doubt. How wise is it to inflict the full brunt of your genius on your colleagues? Will anyone even listen, or be prepared to check your proof?

What you can do is re-format your write-up into a (somewhat longer) PCP proof. Anyone now wishing to verify your argument need only pick a handful of words at random, and follow a set list of instructions to conclude whether it is correct or not. An error might have slipped in on account of faulty working, buggy reformatting, or outright cheating. No matter, it will be caught with overwhelming probability. What the reformatting step does is smear any error all over the proof, making it easier to spot. In much the same way, a diligent sandwich maker will smear a smidgen of jam evenly over his bread, rather than leaving it concentrated in one corner, and so make the whole more savoury.

But before we look under the PCP hood, a word of reassurance to those allergic to maths: the theory extends far beyond mathematical proofs. In its full splendour, PCP asserts that any statement S whose validity can be ascertained by a proof P written over n bits also admits an alternative proof, Q . This proof Q has two appealing features: it can be derived from P in a number of steps proportional to n^c , where c is some constant; and P can be verified by examining only three bits of Q picked at random. If S is true, a correct P will satisfy the verifier with a probability of 99%. If it is not true, any alleged proof P will trigger a rejection from Q with a probability higher than 50% (ref. 4). Not impressed with this error rate? Then all you have to do is pick, instead of three bits of Q , as many bits as are contained in this line of text. The error probability will drop to one in a billion.

Let's take as an example the statement S that graph G is 'three-colourable'. (Thus, if it were a map, it could be coloured in with three colours — say red, green and blue — with no two neighbouring countries sharing the same colour.) The standard proof P of this statement specifies the colour of every vertex (country) in G , and it is verified by checking that no edge (border) is monochromatic. Dinur's formulation³ of PCP, however, instructs the verifier how to build, relatively quickly, a new, error-smearing graph H . This second graph is constructed such that if G is three-colourable, so will H be; if G is not three-colourable, any colouring of H will leave at least a fixed fraction of its edges monochromatic (Fig. 1). In this way, equipped with H , the verifier will be able to probe any alleged colouring at random, with confidence that a monochromatic edge will be spotted if G is not three-colourable.

But how do we efficiently face-lift G into the desired H ? First of all, in a pre-op phase, we inject extra edges into G to make it look random. The surgery itself alternates between two procedures. First, we add edges to make neighbourhoods of G more tightly connected. The intended effect is to amplify the

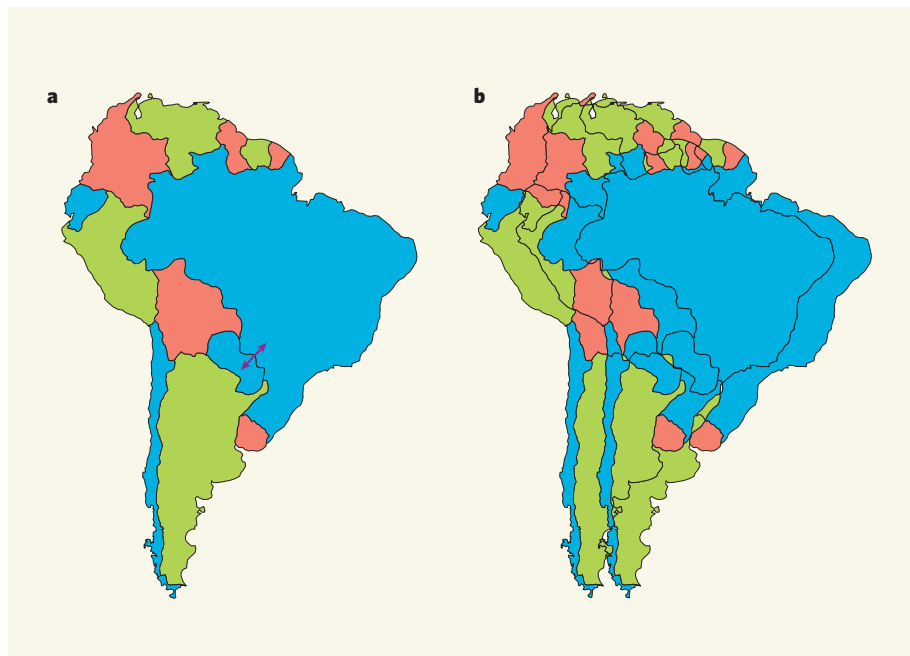


Figure 1 | Probabilistically checkable colouring. The theorem that, on a two-dimensional map, no more than four colours are needed so that no two adjacent countries have the same shading is one of the most notorious in mathematics, and the first to be proved by a computer⁵. Using only three colours is more ticklish, and (a) runs into difficulties even in a finite region of limited boundaries. PCP can be considered as offering an alternative way — rather than laboriously checking every boundary — of ascertaining whether a map claimed to be three-colourable really is. The process is analogous to (b) 'smearing' the map and its error out: in the original, only one error (adjacent countries of the same colour; arrow) occurs, whereas in the smeared map, it is replicated in many areas. Establishing the validity — or not — of the original map with high statistical certainty thus requires the checking of only a small, randomly chosen subregion of the smeared map.

monochromaticism of any faulty colouring. Cosmetic enhancement comes at a price: here, the addition of new colours. So, in a second step, to restore the graph to its original hues of red, green and blue, we call upon error-correcting codes, the redundancy-adding devices invented by coding theorists to make signals immune to noise. This is no coincidence: just as an encoded string with not too many erroneously flipped bits can, through error correction, be restored to its original state, so a PCP proof with not too many errors can be seen to encode a correct proof. Indeed, any error that is not smeared widely enough to be easily spotted is *de facto* inconsequential.

If you think that no one besides children and cartographers has any interest in colouring graphs, think again: the Riemann hypothesis, protein folding, cryptography and most questions in artificial intelligence can be reduced to the three-colourability of some graphs. That universality is another wonder in the computing cosmos.

PCP, and its elementary proof by Dinur³, is the culmination of 40 years of research in the field of computational reduction. Keep in mind that this is all about verifying proofs, not about understanding them — with only three bits! — let alone discovering them. That must still be done the hard way. In the end, PCP boils down to one revolutionary insight: in the art of persuading others of the truth, the ultimate weapon is a set of dice. ■

Bernard Chazelle is in the Department of Computer Science, Princeton University, 35 Olden Street, Princeton, New Jersey 08540-5233, USA.
e-mail: chazelle@cs.princeton.edu

1. Arora, S. & Safra, S. *J. Assoc. Comput. Machin.* **45**, 70–122 (1998).
2. Arora, S., Lund, C., Motwani, R., Sudan, M. & Szegedy, M. *J. Assoc. Comput. Machin.* **45**, 501–555 (1998).
3. Dinur, I. *Proc. 38th Annu. ACM Symp. Theor. Comput.* 241–250 (2006).
4. Håstad, J. *J. Assoc. Comput. Machin.* **48**, 798–859 (2001).
5. Appel, K., Haken, W. & Koch, J. *J. Math.* **21**, 439–567 (1977).

STRUCTURAL BIOLOGY

Dangerous liaisons on neurons

Giampietro Schiavo

Crystal structures show that botulinum toxins bind simultaneously to two sites on neurons. This dual interaction allows them to use a Trojan-horse strategy to enter nerve terminals, with deadly effect.

Botulinum neurotoxins (BoNTs) are some of the most deadly substances known to mankind. By blocking nerve function, they cause botulism, a severe condition that may ultimately lead to muscular and respiratory paralysis. These sophisticated bacterial proteins owe their toxicity to their extraordinary specificity for neurons and to their enzymatic activity. In this issue, papers by Jin *et al.*¹ and Chai *et al.*² describe the mechanisms by which BoNT/B — a toxin that causes human botulism — recognizes the surface of neuron junctions (synapses)*. This work provides insight into how other BoNTs may exert their lethal action, and describes a mode of binding that might be used by other biological compounds.

Once inside a neuron, a single molecule of BoNT is, in principle, capable of deactivating the whole synapse. BoNTs consist of two protein segments, known as the heavy and light chains. It is the light chain that deactivates neuromuscular junctions — the synapses that connect muscles to their controlling neurons — by specifically inhibiting members of the SNARE protein family³. SNARE proteins are distributed over the membranes of all animal and plant cells and are also found on the membranes of synaptic vesicles, the bubble-shaped

organelles that store and release neurotransmitter chemicals at neuron terminals. SNARE proteins are essential for membrane fusion, during which vesicles merge with the cell membrane and release their load. Once the synaptic vesicles have done this, they are recycled by the neuron for further use.

So how do BoNTs enter neurons? The heavy chain is most likely to be responsible. One half of the heavy chain mediates binding to neurons by interacting with lipid molecules (polysialogangliosides, PSGs) in the cell membrane, and with either one of two integral membrane proteins — synaptotagmin I or synaptotagmin II — found in synaptic vesicles. A dual-receptor model for these toxins was proposed long ago⁴, but experimental validation of this theory has required a worldwide effort. The model predicts that the interaction of BoNTs with both PSGs and protein receptors is necessary to explain their awesome potency³, with a different protein receptor being recognized by each BoNT.

Evidence for protein involvement in BoNT binding was scarce until it was discovered⁵ that BoNT/B binds to both PSGs and the part of synaptotagmins that lies inside synaptic vesicles, in the area known as the lumen. More recently, the specific regions of synaptotagmins that bind BoNT/B have been identified⁶,



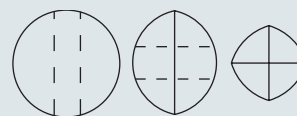
50 YEARS AGO

The recent passion for the progressive splitting up of genera has spread through most branches of zoology and latterly to botany as well. In fact, the present phase in taxonomy is to lump species and split genera... At present, names are very often proposed on flimsy grounds and the work of proving their validity or otherwise is handed down to future workers. New genera are established using merely the specific characters of the unique species included without any additional evidence. The result is that two species such as *Cypraea tigris* L. and *C. pantherina* Sol. are placed in different genera, whereas they are only just distinct species which actually hybridize. Such absurdities should have no standing in nomenclature.

From *Nature* 22 December 1956.

100 YEARS AGO

"Cutting a Round Cake on Scientific Principles"
— Christmas suggests cakes, and with these the wish on my part to describe a method of cutting them that I have recently devised to my own amusement and satisfaction. The problem to be solved was, "given a round tea-cake of some 5 inches across, and two persons of moderate appetite to eat it, in what way should it be cut so as to leave a



minimum of exposed surface to become dry?" The ordinary method of cutting out a wedge is very faulty in this respect... The cuts shown on the figures represent those made with the intention of letting the cake last for three days, each successive operation having removed about one-third of the area of the original disc. A common India-rubber band embraces the whole and keeps its segments together.

From *Nature* 20 December 1906.

50 & 100 YEARS AGO

*This article and the papers concerned^{1,2} were published online on 13 December 2006.

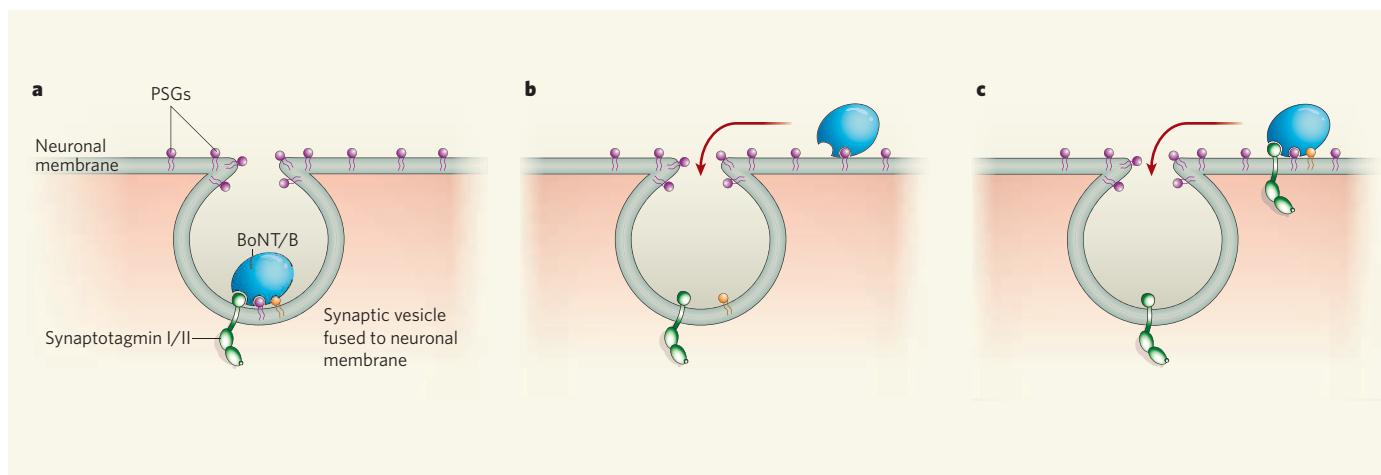


Figure 1 | Possible binding sites of botulinum neurotoxin B (BoNT/B) on neurons. Crystal studies by Jin *et al.*¹ and Chai *et al.*² suggest that BoNT/B invades neurons by stowing away in carriers known as synaptic vesicles. It does this by forming a complex with lipid molecules (polysialogangliosides, PSGs) and with a vesicle protein (either synaptotagmin I or synaptotagmin II) in the neuronal membrane. The complex is stabilized by interactions with neighbouring acidic lipid molecules (orange). BoNT/B could enter open vesicles at the neuron's membrane, in one of three possible sequences. **a**, BoNT/B enters the vesicle directly and forms the required complex. **b**, BoNT/B binds first to PSGs on the membrane, and is then transferred into a synaptic vesicle containing synaptotagmin. **c**, BoNT/B forms a full complex on the membrane, using synaptotagmin left behind by inaccurate vesicle recycling. It is then transferred into the lumen of the vesicle.

as have those bound by the closely related BoNT/G (ref. 7). Taken together, these results suggest that BoNTs enter neurons by stowing away inside recycled synaptic vesicles (Fig. 1), binding simultaneously to PSGs and to the luminal domain of synaptic-vesicle proteins. The toxins then escape from the vesicle lumen when the vesicles are acidified as they reload with neurotransmitters. Another BoNT family member, BoNT/A, has been shown to bind to the synaptic vesicle protein SV2 (refs 8, 9), in agreement with the prediction that each type of BoNT interacts with a different protein.

Jin *et al.*¹ (page 1092) and Chai *et al.*² (page 1096) now present crystal structures showing how BoNT/B binds to the luminal domain of synaptotagmin II. This domain is unstructured in solution, but upon binding to the heavy chain of BoNT/B it folds into a helix, which then enters a cleft adjacent to the PSG-binding site of the toxin. The amino acids lining this cleft are very similar to those found in BoNT/G (ref. 10), but differ in the other toxin family members, which explains why different BoNTs recognize distinct protein receptors. Mutations in the synaptotagmin-binding cleft greatly reduce the toxicity of BoNT/B, more so than mutations in the PSG-binding pocket. Binding to synaptotagmin therefore seems to be central to BoNT/B toxicity, so blocking this interaction looks like a good strategy for preventing botulism.

Although these studies^{1,2} were performed in the absence of vesicle membranes, the binding of PSGs and synaptotagmin II to adjacent sites on BoNT/B offers clues as to how the toxin might interact with these membranes. The two binding sites would firmly anchor the tip of BoNT/B to the vesicle's inner surface, constraining the toxin's mobility. The rigid character of this interaction might be further enhanced by the association of the

toxin's heavy chain with nearby negatively charged lipid molecules, which play an accessory role in stabilizing the toxin on membranes³.

A difference between the binding modes proposed by Jin *et al.*¹ and Chai *et al.*² concerns the orientation of the remaining portion of the BoNT heavy chain. This section has two long helices that are thought to insert into the vesicle membrane, so helping the light chain to reach the cytoplasm. Jin and colleagues suggest that these helices sit orthogonal to the plane of the membrane, where they resemble a plunger. But Chai *et al.* propose a parallel orientation, which would place hydrophobic areas of the helices close to the inner surface of the vesicle, facilitating their insertion into the membrane. These two spatial arrangements could be snapshots of the journey that BoNT/B takes when close to the membrane of the neuron.

Another issue arising from this work concerns the partial overlap of the two binding sites. Chai *et al.* argue that complex PSGs could contact both BoNT/B and synaptotagmin. This mode of binding would lock the toxin to the neuronal surface, and could explain why BoNTs have a low tendency to detach from the membrane once in place. Furthermore, these studies^{1,2} do not define the sequence of events that leads to the membrane-binding of BoNT/B (Fig. 1). The simplest possibility is that BoNT/B binds to PSGs and synaptotagmin within the lumen of a synaptic vesicle that is fused to the neuron membrane. However, the recruitment of BoNT/B to the membrane could start outside the vesicle, if it binds first to PSGs and then transfers to a recycling vesicle that contains synaptotagmin. Alternatively, upon PSG binding, BoNT/B might interact with the fraction of synaptotagmin that is present on the surface of the neuron as a result

of inaccurate synaptic-vesicle recycling¹¹.

To identify the most relevant mechanism, it is essential to assess whether PSGs are present on the luminal side of the synaptic-vesicle membrane, but conflicting data have been reported on this issue. Interestingly, tetanus toxin (TeNT) — which is closely related to BoNTs and has a similar mechanism of action — also requires PSGs for its neuronal binding. However, PSGs are excluded from the lumen of the organelles that allow TeNT to enter neuromuscular junctions¹². This casts doubt on whether PSGs play a role in BoNT binding inside synaptic vesicles.

Such gaps in our knowledge serve only to heighten our interest in these fascinating neurotoxins. BoNTs are powerful tools for biologists and find widespread use as therapeutics for the treatment of certain nervous-system diseases. For these reasons, the papers^{1,2} reported here are of tremendous value.

Giampietro Schiavo is at the Molecular Neuropathobiology Laboratory, Cancer Research UK, London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.
e-mail: giampietro.schiavo@cancer.org.uk

1. Jin, R., Rummel, A., Binz, T. & Brunger, A. T. *Nature* **444**, 1092–1095 (2006).
2. Chai, Q. *et al.* *Nature* **444**, 1096–1100 (2006).
3. Schiavo, G., Matteoli, M. & Montecucco, C. *Physiol. Rev.* **80**, 717–766 (2000).
4. Montecucco, C. *Trends Biochem. Sci.* **11**, 314–317 (1986).
5. Nishiki, T. *et al.* *FEBS Lett.* **378**, 253–257 (1996).
6. Dong, M. *et al.* *J. Cell Biol.* **162**, 1293–1303 (2003).
7. Rummel, A., Karnath, T., Henke, T., Bigalke, H. & Binz, T. *J. Biol. Chem.* **279**, 30865–30870 (2004).
8. Dong, M. *et al.* *Science* **312**, 592–596 (2006).
9. Mahrhold, S., Rummel, A., Bigalke, H., Davletov, B. & Binz, T. *FEBS Lett.* **580**, 2011–2014 (2006).
10. Rummel, A. *et al.* *Proc. Natl Acad. Sci. USA* (in the press).
11. Fernandez-Alfonso, T., Kwan, R. & Ryan, T. A. *Neuron* **51**, 149–151 (2006).
12. Deinhardt, K., Berninghausen, O., Willison, H. J., Hopkins, C. R. & Schiavo, G. *J. Cell Biol.* **174**, 459–471 (2006).

BRIEF COMMUNICATIONS

Parthenogenesis in Komodo dragons

Should males and females be kept together to avoid triggering virgin birth in these endangered reptiles?

Parthenogenesis, the production of offspring without fertilization by a male, is rare in vertebrate species, which usually reproduce after fusion of male and female gametes. Here we use genetic fingerprinting to identify parthenogenetic offspring produced by two female Komodo dragons (*Varanus komodoensis*) that had been kept at separate institutions and isolated from males; one of these females subsequently produced additional offspring sexually. This reproductive plasticity indicates that female Komodo dragons may switch between asexual and sexual reproduction, depending on the availability of a mate — a finding that has implications for the breeding of this threatened species in captivity. Most zoos keep only females, with males being moved between zoos for mating, but perhaps they should be kept together to avoid triggering parthenogenesis and thereby decreasing genetic diversity.

Komodo dragons, the largest of the lizards, are under threat¹ as wild populations become smaller and more fragmented, as are 341 other species of reptile. At least 52 zoological institutions are cooperating in a successful international breeding programme with these lizards, but until now parthenogenesis has never been reported.

There are only two sexually mature female Komodo dragons in Europe, both of which were bred in captivity and are crucial to the success of the regional breeding programme. One of these (Flora, at Chester Zoo, UK) has never been kept with a male but has nevertheless produced a clutch of 25 eggs, of which 11 seemed to be viable. Three of these eggs collapsed early during incubation and provided embryonic material for genetic fingerprinting. The remaining eight eggs are developing normally and are expected to hatch in January 2007. Another captive-bred female (Sungai, at London Zoo, UK), now deceased, produced four viable eggs (from a clutch of 22) 2.5 years after her last contact with a male (Kimaan), which could be explained either by long-term sperm storage or by parthenogenesis; the eggs hatched 7.5 months later, and the young seem to be healthy (Fig. 1). Sungai subsequently laid a second clutch of six eggs 2 months after mating with a different male (Raja), of which just one hatched (for a full history of these animals, see supplementary information).

We analysed the parentage of the eggs and offspring by genetic fingerprinting² (see

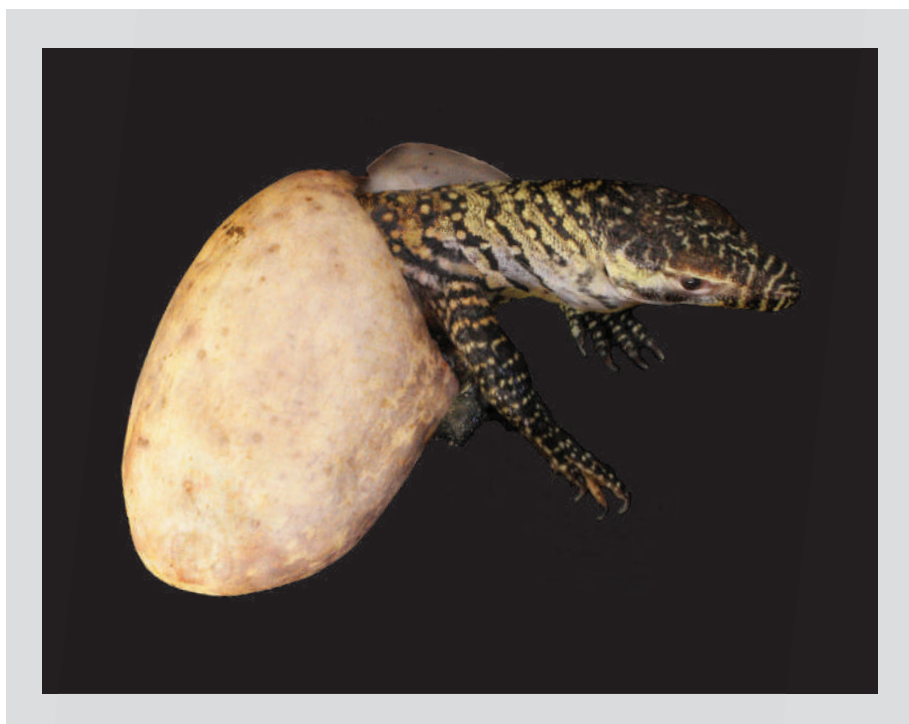


Figure 1 | Hatching of a Komodo dragon (*Varanus komodoensis*) from an egg produced asexually.

supplementary information). In the clutches of both females, we found that all offspring produced in the absence of males were parthenogens: the overall combined clutch genotype reconstructed that of their mother exactly. Although all offspring were homozygous at all loci, they were not identical clones. Parthenogenesis was therefore confirmed by exclusion (clutches had different alleles from potential fathers) and by the fact that the probability of obtaining a clutch of homozygous individuals after sexual reproduction was very low ($P < 0.0001$). Sungai's resumption of sexual reproduction confirmed that parthenogenesis was not a fixed reproductive trait (that is, it is facultative) and that asexual reproduction is likely to occur only when necessary.

Parthenogenesis has been reported in about 70 vertebrate species (roughly 0.1%)^{3,4}. Our demonstration of facultative parthenogenesis in Komodo dragons is unexpected in such large reptiles. It occurs in captive snakes⁵ and has been implicated in one other species of Argus monitor lizard (*Varanus panoptes*)⁶. Our observations of two separate occurrences of parthenogenesis at two different institutions

indicate that this reproductive strategy might not be so unusual when Komodo dragons are isolated, even though reproductive plasticity in species thought to reproduce sexually is unexpected owing to the strict requirement to maintain diploidy.

Parthenogenesis presents a previously unrecognized problem for the genetic management of threatened populations. Although captive breeding can be an essential part of a species' conservation, our results indicate that studbook records of reptile species might no longer be accurate. A pressing concern with parthenogenesis is instantaneous homozygosity of the entire genome, as this inbreeding carries an associated risk of reduced fitness and an increased probability of extinction^{7,8}. It is noteworthy, however, that the four Komodo dragon parthenogens at London Zoo are healthy and growing and behaving normally. It is common practice to keep extra females without males in captivity to maintain a sex-ratio bias towards the reproductively limiting sex and, because these are solitary animals, to reduce the risk of aggressive interactions. However, they are then subjected to strong selective pressures — as experienced by island

I. STEPHEN

colonists of a sexually reproducing damselfly (*Ischnura hastata*), for example, that became exclusively parthenogenetic⁹. Parthenogenesis can also bias the sex ratio: in *Varanus* species, females have dissimilar chromosomes (Z and W), whereas the combination ZZ produces males¹⁰, so the parthenogenetic mechanism can produce only homozygous (ZZ or WW) individuals and therefore no females.

Parthenogenesis in wild Komodo dragons could be adaptive, given that viable offspring are always male and that sexual reproduction can resume, albeit between related individuals, in a colony founded by a single unfertilized female. Fewer than 4,000 Komodo dragons remain in the wild, of which perhaps fewer than 1,000 are mature females (T. Jessop, personal communication). Our discovery of the potential for asexual reproduction in this species, and possibly in other reptiles presumed

until now to be exclusively sexual, calls for further investigation into the genetic load experienced by the parthenogens, the frequency with which asexual offspring occur in captive and in wild populations, and the fitness consequences associated with facultative parthenogenesis.

Phillip C. Watts*, **Kevin R. Buley†**,
Stephanie Sanderson‡, **Wayne Boardman‡**,
Claudio Ciofi§, **Richard Gibson‡**

*School of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, UK
e-mail: p.c.watts@liv.ac.uk

†North of England Zoological Society, Chester Zoo, Upton-by-Chester CH2 1LH, UK

‡Zoological Society of London, Regents Park, London NW1 4RY, UK

§Department of Animal Biology and Genetics, University of Florence, 50125 Florence, Italy

1. Ciofi, C. & De Boer, M. *Herpetol. J.* **14**, 99–107 (2004).

2. Ciofi, C. & Bruford, M. W. *Mol. Ecol.* **7**, 134–136 (1998).
3. White, M. J. D. *Animal Cytology and Evolution* (Cambridge University Press, Cambridge, UK, 1973).
4. Avise, J. C., Quattro, J. M. & Vrijenhoek, R. C. in *Evolutionary Biology* (eds Hecht, M. K., Wallace, B. & MacIntyre, R. J.) 225–246 (Plenum, New York, 1992).
5. Groot, T. V. M., Bruins, E. & Breeuwer, J. A. *Heredity* **90**, 130–135 (2003).
6. Lenk, P., Eidenmueller, B., Staudter, H., Wicker, R. & Wink, M. *Amphibia-Reptilia* **26**, 507–514 (2005).
7. Saccheri, I. J. et al. *Nature* **392**, 491–494 (1998).
8. Spielman, D., Brook, B. W. & Frankham, R. *Proc. Natl Acad. Sci. USA* **101**, 15261–15264 (2004).
9. Sherratt, T. N. & Beatty, C. D. *Nature* **435**, 1039–1040 (2005).
10. Halverson, J. & Spelman, L. H. in *Biology and Conservation of Komodo Dragons* (eds Murphy, J. B., Ciofi, C., de La Panouse, C. & Walsh, T.) 165–177 (Smithsonian Institution, Washington DC, 2002).

Supplementary information accompanies this communication on Nature's website.

Received 4 October; accepted 16 November 2006.

Competing financial interests: declared none.

doi:10.1038/nature4441021a

MICROBIAL ECOLOGY

Human gut microbes associated with obesity

Two groups of beneficial bacteria are dominant in the human gut, the Bacteroidetes and the Firmicutes. Here we show that the relative proportion of Bacteroidetes is decreased in obese people by comparison with lean people, and that this proportion increases with weight loss on two types of low-calorie diet. Our findings indicate that obesity has a microbial component, which might have potential therapeutic implications.

Trillions of microbes live in the human gut, helping to break down otherwise indigestible foods¹. Transplanting the gut microbiota from normal mice into germ-free recipients increases their body fat without any increase in food consumption², raising the possibility that the composition of the microbial community in the gut affects the amount of energy extracted from the diet².

The relative abundance of the two predominant bacterial divisions (deep evolutionary lineages or superkingdoms) in mice differs between lean and obese animals: mice that are genetically obese (*ob/ob*) have 50% fewer Bacteroidetes, and correspondingly more Firmicutes, than their lean (+/+) siblings³. In an accompanying Article⁴, we show that the gut microbiota in these *ob/ob* mice are more effective at releasing calories from food during digestion than are the +/+ microbiota: this trait is transmissible to germ-free recipients, resulting in greater adiposity.

To investigate the relation between gut microbial ecology and body fat in humans, we studied 12 obese people, who were randomly assigned to either a fat-restricted (FAT-R) or to a carbohydrate-restricted (CARB-R) low-

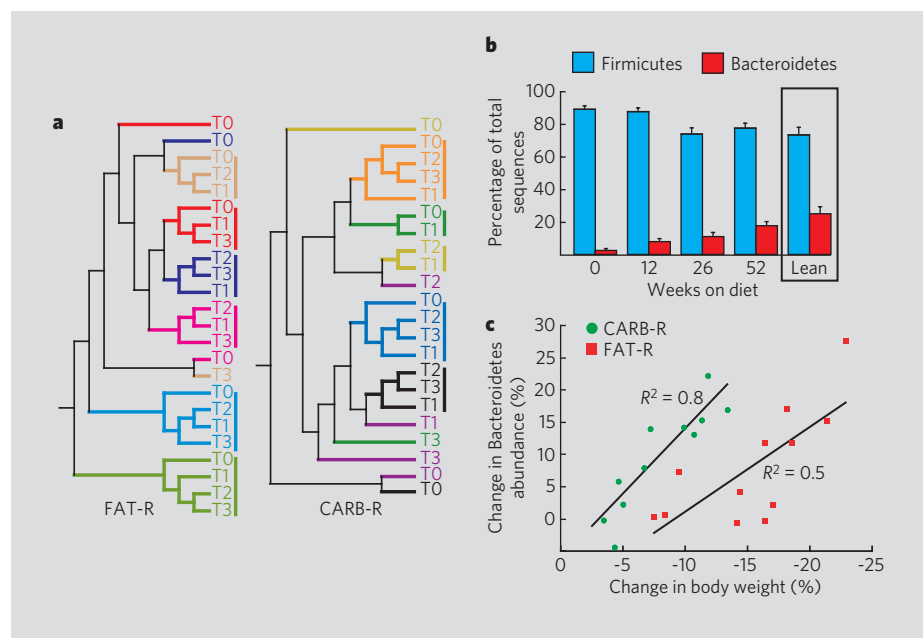


Figure 1 | Correlation between body-weight loss and gut microbial ecology. **a**, Clustering of 16S ribosomal RNA gene sequence libraries of faecal microbiota for each person (in different colours) and time point in diet therapy (T0, baseline; T1, 12 weeks; T2, 26 weeks; T3, 52 weeks) in the two diet-treatment groups (fat restricted, FAT-R; carbohydrate restricted, CARB-R), based on UniFrac analysis of the 18,348-sequence phylogenetic tree. **b**, Relative abundance of Bacteroidetes and Firmicutes. For each time point, values from all available samples were averaged (n was 11 or 12 per time point). Lean-subject controls include four stool samples from two people taken 1 year apart, plus three other stool samples⁶. Mean values $\pm$ s.e. are plotted. **c**, Change in relative abundance of Bacteroidetes in subjects with weight loss above a threshold of 2% weight loss for the CARB-R diet and 6% for the FAT-R diet.

calorie diet. The composition of their gut microbiota was monitored over the course of 1 year by sequencing 16S ribosomal RNA genes from stool samples (for details, see supplementary information).

The resulting data set of 18,348 bacterial 16S rRNA sequences revealed that most (70%) of the 4,074 identified species-level phylogenetic types (phylotypes) were unique to each person (see supplementary information). Despite the

marked interpersonal differences in species-level diversity, members of the Bacteroidetes and Firmicutes divisions dominated the microbiota (92.6% of all 16S rRNA sequences).

Bacterial lineages were remarkably constant within people over time: communities from the same person were generally more similar to one another than to those from other people (Fig. 1a). Before diet therapy, obese people had fewer Bacteroidetes ($P < 0.001$) and more Firmicutes ($P = 0.002$) than did lean controls (Fig. 1b). Over time, the relative abundance of Bacteroidetes increased ($P < 0.001$) and the abundance of Firmicutes decreased ($P = 0.002$), irrespective of diet type (Fig. 1b).

This change was division-wide and not due to blooms or extinctions of specific bacterial species: bacterial diversity remained constant over time. Increased abundance of Bacteroidetes correlated with percentage loss of body weight (R^2 was 0.8 for the CARB-R diet and 0.5 for the FAT-R diet, $P < 0.05$; Fig. 1c), and not with changes in dietary calorie content

over time (R^2 was 0.06 for the CARB-R diet and 0.09 for the FAT-R diet). This correlation held only after the person had lost at least 6% of their body weight on the FAT-R diet and at least 2% on the CARB-R diet.

Obesity is, to our knowledge, the only condition in which a pronounced, division-wide change in microbial ecology is associated with host pathology. The factors that drive shifts in representation at such broad taxonomic levels must operate on highly conserved bacterial traits as they are shared by a great variety of phylotypes within the divisions. The gut habitat itself selects for specific ratios of divisions: microbiota transplanted to germ-free recipients of a different species reconfigure to match the community structure usually present in the recipient⁵. The coexistence of Bacteroidetes and Firmicutes in the gut implies minimized competition for resources through cooperation or specialization: the obese gut has as-yet uncharacterized properties that tip the balance towards the Firmicutes.

The dynamic linkage between adiposity and gut microbial ecology described here, together with our results from mice⁴, indicates that manipulation of gut microbial communities could be another approach in the treatment of obesity.

Ruth E. Ley, Peter J. Turnbaugh, Samuel Klein, Jeffrey I. Gordon

Washington University School of Medicine,
St Louis, Missouri 63108, USA

e-mail: jgordon@wustl.edu

1. Gill, S. R. *et al. Science* **312**, 1355–1359 (2006).
2. Backhed, F. *et al. Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
3. Ley, R. E. *et al. Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
4. Turnbaugh, P. J. *et al. Nature* **444**, 1027–1031 (2006).
5. Rawls, J. F., Mahowald, M. A., Ley, R. E. & Gordon, J. I. *Cell* **127**, 423–433 (2006).
6. Eckburg, P. B. *et al. Science* **308**, 1635–1638 (2005).

Supplementary information accompanies this communication on Nature's website.

Received 8 October; accepted 10 November 2006.

Competing financial interests: declared none.

doi:10.1038/nature4441022a

BIOMECHANICS

Rubber bands reduce the cost of carrying loads

Vertical movement of the hip during locomotion causes a loaded backpack to be accelerated with each step¹, which imposes large peak forces on the wearer. Here we show that using bungee cords to suspend the load from a backpack frame reduces not only its vertical movement, and hence its vertical force on the carrier, but also the energetic cost of walking with the pack. This permits larger loads to be carried while moving rapidly, and at the same time reduces the risk of orthopaedic and muscular injury.

Backpacks have an inherent limitation: because the hips move up and down by 5–7 cm during walking², the mass of a backpack — which is usually attached tightly to the body — must undergo a similar vertical displacement^{1,3}. As a result, the peak vertical forces acting on the body increase by up to twice those imposed by the static weight, owing to the acceleration of the added mass.

If the load were to stay positioned at the same height during walking and running, it would not exert any accelerative force on each step — a strategy that has been exploited by Asian merchants running with springy bamboo poles¹. We designed a backpack in which the load is suspended from an external frame by a compliant coupling and is therefore largely uncoupled from the body's movements. A locking mechanism enables the coupling between the load and the body to be altered from 'suspended' to 'locked' (Fig. 1).

While walking at 5.6 km h⁻¹, the vertical oscillation of a 27-kg load is reduced from

68.5 mm in the locked backpack (Fig. 2a) to 26.5 mm in the suspended backpack ($P < 0.001$, paired t -test, $n = 6$; see table in supplementary information). The oscillation is reduced mainly because the amount of movement of the load relative to the frame is similar in magnitude and nearly 180° out of phase with the frame's displacement relative to the ground. The reduction causes an 82% decrease in peak accelerative vertical force (Fig. 2b), which declines from 185 to 30.5 N in the locked versus the suspended backpack ($P < 0.001$, paired t -test, $n = 6$; see table in supplementary information). The total peak vertical force drops by 33%.

The metabolic cost of walking with the load falls from 640 W for the locked backpack to 600 W ($P = 0.027$, paired t -test, $n = 6$; see table in supplementary information) for the suspended backpack, because the force exerted by the load is reduced during the energetically expensive 'double-support' phase of walking (when both feet are on the ground)^{4,5}. Although the average force (that is, weight or static force) from the load is the same for the locked and suspended backpack, the locked backpack concentrates the forces in the double-support phase (Fig. 2b), probably resulting in greater ground-reaction force and loss of mechanical energy. The consequent reduction in force during the energy-conserving inverted-pendulum phase (when only one foot is on the ground) yields little energetic benefit.

Although the reduction in metabolism is modest (6.2%) in terms of the total metabolic

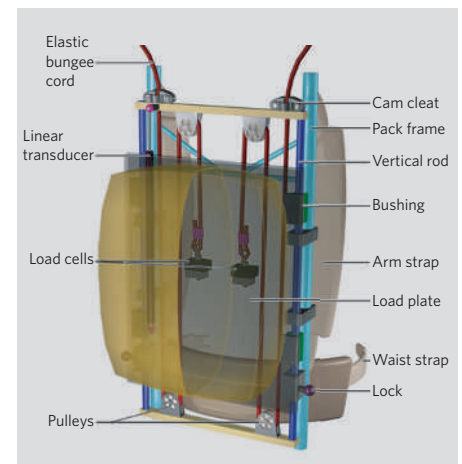


Figure 1 | Suspended-load ergonomic backpack. The pack frame is fixed to the body but the load, mounted on the load plate, is suspended by elastic bungee cord(s) from the frame (shown in light blue and gold). The tension in the elastic cord is set by cam cleats on the top of the frame. When the backpack's locking mechanism is disengaged, the suspended load is free to ride up and down on bushings constrained to vertical rods during walking or running; locking prevents movement and the backpack behaves like a normal rigid backpack. For movies, see supplementary information.

rate, it represents 23% of the extra metabolic power (176 W) that is required to walk with a 27-kg load rather than with an empty backpack. Walking with 27 kg in the suspended

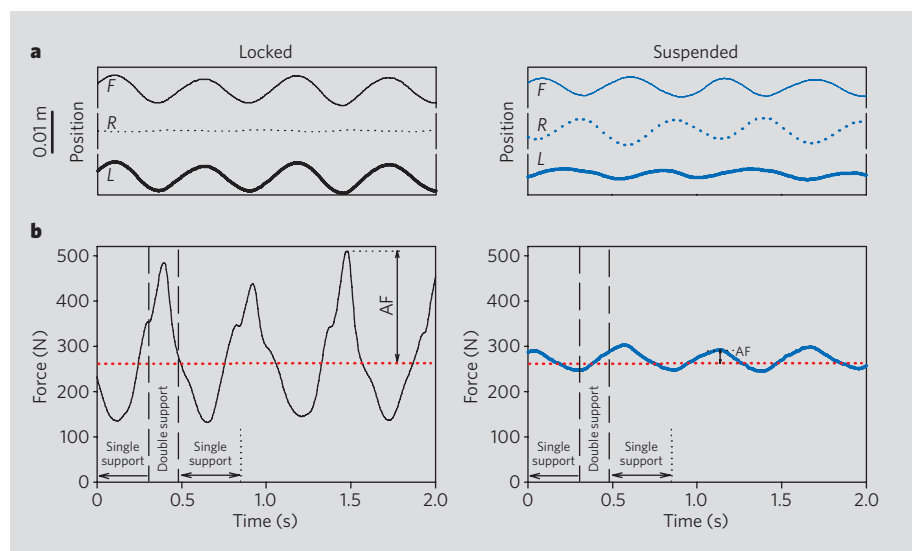


Figure 2 | Vertical position of, and force exerted by, a 27-kg load carried by a hiker walking at 5.6 km h⁻¹. **a**, The displacement of the load (*L*) with the locked backpack (left panel) is equal to the displacement of the frame (*F*) plus the displacement of the load relative to the frame (*R*). Vertical movement of the load (*L*) with the suspended backpack (right panel) is reduced because the load moves down in the frame (*R*) to compensate for the upward movement of the frame (*F*). **b**, The large vertical displacement of the locked backpack (left) requires large accelerative forces (AF) — the difference between the total vertical force and the weight (red dotted line) to be exerted on the load by the frame; the load, in turn, exerts large forces back on the walker. With the suspended backpack (right), therefore, little accelerative force is required. The double-support and single-support ('inverted pendulum') phases of walking are shown for the first step.

backpack is equivalent to walking with 21.7 kg in the locked backpack in terms of metabolic cost, so — for a given metabolic rate — the suspended backpack enables a substantially heavier load to be carried (a mean of 5.3 kg $\pm$ 3.4 kg more, $n = 6$; see table in supplementary information). These findings show that a passive device can make the energetic cost of carrying loads more economical^{1,6,7}.

The large weight of backpacks carried by children is internationally recognized as a public health problem^{8–10}. Lower peak vertical forces during load carrying could reduce muscle and orthopaedic injury. Total peak vertical forces are even more harmful during running, reaching more than three times the static force of the load. The suspended backpack cuts the vertical oscillation of the load from 77.8

to 26.0 mm ($P < 0.001$, paired t -test, $n = 5$; see table in supplementary information), reducing accelerative forces by 86% (s.d. $\pm$ 8%, $n = 5$) and total peak vertical force by 60%, thereby greatly easing discomfort. Being able to run with this backpack, rather than being forced to walk, would be particularly useful to emergency personnel who need to carry equipment rapidly to disaster sites.

Lawrence C. Rome*†, **Louis Flynn***,
Tae-seung D. Yoo*

*Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

e-mail: lrome@sas.upenn.edu

†Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

1. Kram, R. *J. Appl. Physiol.* **71**, 1119–1122 (1991).
2. Gard, S. A., Miff, S. C. & Kuo, A. D. *Hum. Mov. Sci.* **22**, 597–610 (2004).
3. Rome, L. C., Flynn, L., Goldman, E. M. & Yoo, T. D. *Science* **309**, 1725–1728 (2005).
4. Grabowski, A., Farley, C. T. & Kram, R. *J. Appl. Physiol.* **98**, 579–583 (2005).
5. Donelan, J. M., Kram, R. & Kuo, A. D. *J. Exp. Biol.* **205**, 3717–3727 (2002).
6. Balogun, J. A. et al. *Ergonomics* **29**, 1623–1635 (1986).
7. Chatterjee, A., Pratap, R., Reddy, C. K. & Ruina, A. *Int. J. Robot. Res.* **21**, 621–34 (2002).
8. Troussier, B., Davoine, P., de Gaudemaris, R., Fauconnier, J. & Phelip, X. *Scand. J. Rehabil. Med.* **26**, 143–146 (1994).
9. Negrini, S. & Carabalona, R. *Spine* **27**, 187–195 (2002).
10. Taimela, S., Kujala, U. M., Salminen, J. J. & Viljanen, T. *Spine* **22**, 1132–1136 (1997).

Supplementary information accompanies this communication of Nature's website.

Received 21 September; accepted 22 November 2006.

Competing financial interests: declared (see online communication for details).

doi: 10.1038/nature4441023a

OLFACTION

Underwater 'sniffing' by semi-aquatic mammals

Terrestrial species that forage underwater face challenges because their body parts and senses are adapted for land — for example, it is widely held that mammals cannot use olfaction underwater because it is impossible for them to inspire air (sniff) to convey odorants to the olfactory epithelium^{1–5}. Here I describe a mechanism for underwater sniffing used by the semi-aquatic star-nosed mole (*Condylura cristata*) and water shrew (*Sorex palustris*). While underwater, both species exhale air bubbles onto objects or scent trails and then re-inspire the bubbles to carry the smell back through the nose. This newly described behaviour provides a mechanism for mammalian olfaction underwater.

High-speed video recordings of star-nosed moles reveal that they continuously emit and re-inhale air from their nostrils while foraging underwater (Fig. 1; for videos, see sup-

plementary information), indicating that they could be 'sniffing' odours while submerged. To test this idea, moles were trained to follow an underwater scent trail that was randomly laid on either of two paths leading to food (for details of methods, see supplementary information). Trails were laid in a channel covered with a steel grid that allowed the air bubble to pass through it, contact the scent trail, and be re-inhaled, while at the same time preventing contact with the mole's star nose (see video in supplementary information).

The five moles tested on an earthworm scent followed the underwater scent trails to reach a reward with an average accuracy of 85% (ranging from 75% to 100% correct for the 20 trials used for each mole; Fig. 2a), and the two moles tested on a fish scent followed the smell with 85% and 100% accuracy, respectively (see supplementary information).



Figure 1 | Star nose of the mole (*Condylura cristata*) breathing air while underwater. Bubbles are of comparable size to 'sniff' volumes above water in small mammals).

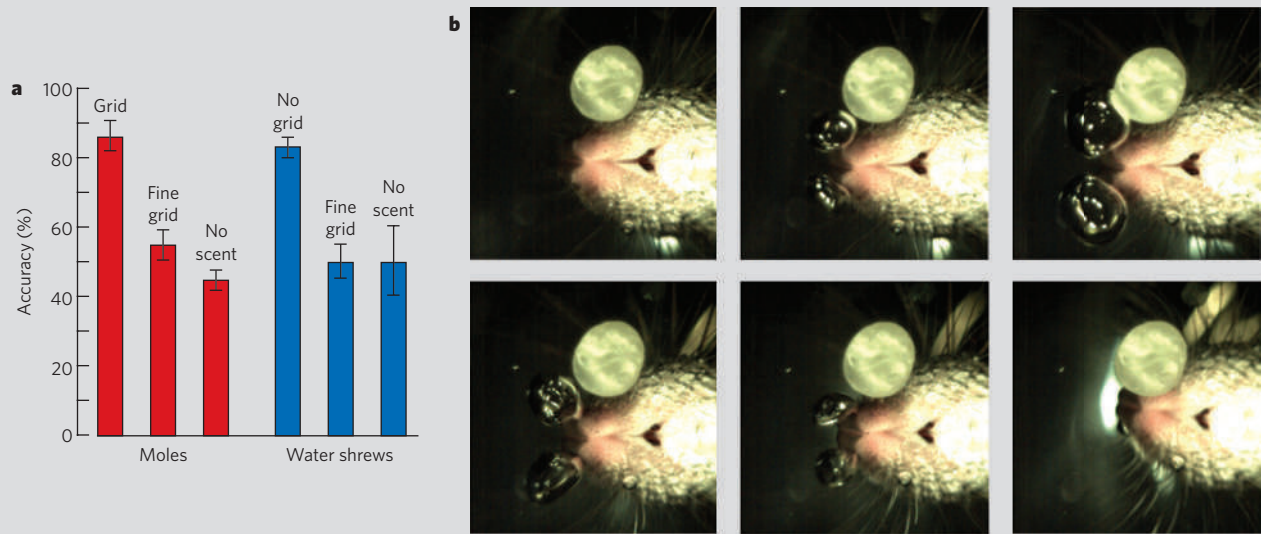


Figure 2 | Underwater olfaction in semi-aquatic moles and shrews. **a**, Results of tracking of scent trails (20 trials per experiment) by the star-nosed mole (*Condylura cristata*) and by the water shrew (*Sorex palustris*). Performance was significant ($P \leq 0.05$ s.e.m. is indicated for each bar) for moles (red bars) sniffing through a grid ($n = 5$) and for water shrews sniffing in the absence of a grid (blue bars, $n = 2$). Performance was at chance when bubbles were blocked with a fine grid ($n = 3$), or when the trail had no scent. **b**, Underwater sniffing by two water shrews. The bubble is in contact with a wax object (green). Top panels, expiration; lower panels, inspiration.

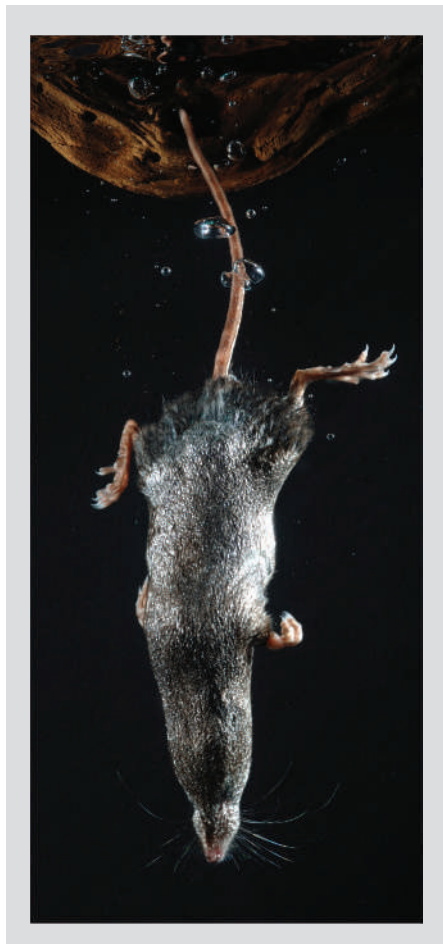


Figure 3 | A diving American water shrew (*Sorex palustris*). This semi-aquatic species has water-repellent fur and is the smallest mammalian diver.

Three moles were also tested with a grid of finer mesh that prevented bubbles from contacting the scent trail. Their performance dropped to chance (that is, 50% in a two-choice test), suggesting that close proximity or contact of the bubbles with the odorants is important in this behaviour; performance was also at chance for two moles when there was no scent trail to follow (Fig. 2a).

Star-nosed moles therefore seem to have adapted their olfactory system for use underwater. To test whether other semi-aquatic mammals might show the same behaviour, I investigated whether the water shrew (Fig. 3) uses a similar strategy. Four water shrews were each seen to emit and re-inhale bubbles that were frequently in contact with objects before being re-inhaled (Fig. 2b, and see video in supplementary information).

Two water shrews were trained to follow an underwater fish-scent trail. No grid was used to prevent direct contact with the scent, because the nose of these animals does not physically permit sniffing through a grid; they depend on their whiskers, or vibrissae⁶, for exploratory contact, unlike the mole, which uses its sensitive star nose. One shrew was accurate in 80% of trials and the other in 85% of trials, and both performed at chance when a blocking grid or a no-scent trail was used (Fig. 2a). These findings show that water shrews are also able to use olfaction underwater.

Can the observed behaviour be classified as 'sniffing' underwater? The volumes of air that are expired and re-inspired (0.06–0.10 ml), the rate of airflow (5–6 ml s⁻¹), the frequency (8–12 Hz) and the context (while exploring an object or pausing) were all remarkably simi-

lar to sniffing, as described in air^{7,8}. Rats, for example, sniff at a similar frequency (4–12 Hz) and airflow rate (5 ml s⁻¹) and, corrected for body weight, by using comparable volumes of air (0.25 ml)⁹.

These results call for a reassessment of the assumption that olfaction is useless underwater and raise the possibility that air could be an intermediate substrate for odorant transport in other aquatic animals as well.

Kenneth C. Catania

Department of Biological Sciences, Vanderbilt University, VU Station B, Box 35-1634, Nashville, Tennessee 37235, USA
e-mail: ken.catania@vanderbilt.edu

1. Howell, A. B. *Aquatic Mammals: Their Adaptations to Life in the Water* (Thomas, Springfield, Illinois, 1930).
2. Stephan, H., Baron, G. & Fons, R. Z. *Zool. Syst. Evol.* **22**, 328–342 (1984).
3. Estes, J. A. in *Carnivore Behavior, Ecology, and Evolution* vol. 1 (ed. Gittleman, J. L.) 242–282 (Cornell University Press, Ithaca, New York, 1989).
4. Sanchez-Villagra, M. R. & Asher, R. J. *Mammalia* **66**, 93–109 (2002).
5. Repenning, C. A. *Syst. Zool.* **25**, 375–390 (1976).
6. Catania, K. C. *Brain. Behav. Evol.* **56**, 146–174 (2000).
7. Kepecs, A., Uchida, N. & Mainen, Z. F. *Chem. Senses* **31**, 167–179 (2006).
8. Uchida, N., Kepecs, A. & Mainen, Z. F. *Nature Rev. Neurosci.* **7**, 485–491 (2006).
9. Youngentob, S. L., Mozell, M. M., Sheehee, P. R. & Hornung, D. E. *Physiol. Behav.* **41**, 59–69 (1987).

Supplementary information accompanies this communication on Nature's website.

Received 18 August; accepted 28 November 2006.

Competing financial interests: none declared.
doi:10.1038/nature4441024a

An obesity-associated gut microbiome with increased capacity for energy harvest

Peter J. Turnbaugh¹, Ruth E. Ley¹, Michael A. Mahowald¹, Vincent Magrini², Elaine R. Mardis^{1,2} & Jeffrey I. Gordon¹

The worldwide obesity epidemic is stimulating efforts to identify host and environmental factors that affect energy balance. Comparisons of the distal gut microbiota of genetically obese mice and their lean littermates, as well as those of obese and lean human volunteers have revealed that obesity is associated with changes in the relative abundance of the two dominant bacterial divisions, the Bacteroidetes and the Firmicutes. Here we demonstrate through metagenomic and biochemical analyses that these changes affect the metabolic potential of the mouse gut microbiota. Our results indicate that the obese microbiome has an increased capacity to harvest energy from the diet. Furthermore, this trait is transmissible: colonization of germ-free mice with an 'obese microbiota' results in a significantly greater increase in total body fat than colonization with a 'lean microbiota'. These results identify the gut microbiota as an additional contributing factor to the pathophysiology of obesity.

The human 'metagenome' is a composite of *Homo sapiens* genes and genes present in the genomes of the trillions of microbes that colonize our adult bodies. The latter genes are thought to outnumber the former by several orders of magnitude¹. 'Our' microbial genomes (the microbiome) encode metabolic capacities that we have not had to evolve wholly on our own^{2,3}, but remain largely unexplored. These include degradation of otherwise indigestible components of our diet⁴, and therefore may have an impact on our energy balance.

Colonization of adult germ-free mice with a distal gut microbial community harvested from conventionally raised mice produces a dramatic increase in body fat within 10–14 days, despite an associated decrease in food consumption⁵. This change involves several linked mechanisms: microbial fermentation of dietary polysaccharides that cannot be digested by the host; subsequent intestinal absorption of monosaccharides and short-chain fatty acids; their conversion to more complex lipids in the liver; and microbial regulation of host genes that promote deposition of the lipids in adipocytes⁵. These findings have led us to propose that the microbiota of obese individuals may be more efficient at extracting energy from a given diet than the microbiota of lean individuals^{2,5}.

In a previous study, we performed a comparative 16S-rRNA-gene-sequence-based survey of the distal gut microbiota of adult C57BL/6J mice homozygous for a mutation in the leptin gene (*Lep^{ob}*) that produces obesity, as well as the microbiota of their lean (*ob/+* and *+/+*) littermates⁶. Members of two of the 70 known divisions of Bacteria^{7,8}, the Bacteroidetes and the Firmicutes, consisted of more than 90% of all phylogenetic types in both groups of mice, just as they do in humans^{6,9,10}. However, the relative abundance of the Bacteroidetes in *ob/ob* mice was lower by 50%, whereas the Firmicutes were higher by a corresponding degree⁶. These differences were division-wide, and not attributable to differences in food consumption (a runted *ob/ob* mouse weighed less than his *ob/ob* littermates owing to reduced chow consumption, but still exhibited a markedly greater per cent body fat and ratio of Firmicutes to Bacteroidetes)⁶.

We have observed analogous differences in the distal gut microbiota of obese versus lean humans; the relative abundance of Bacteroidetes increases as obese individuals lose weight on either a fat- or a carbohydrate-restricted low-calorie diet. Moreover, the

increase in Bacteroidetes was significantly correlated to weight loss but not to total caloric intake⁹.

To determine if microbial community gene content correlates with, and is a potential contributing factor to obesity, we characterized the distal gut microbiomes of *ob/ob*, *ob/+*, and *+/+* littermates by random shotgun sequencing of their caecal microbial DNA. Mice were used for these comparative metagenomics studies to eliminate many of the confounding variables (environment, diet and genotype) that would make such a proof of principle experiment more difficult to perform and interpret in humans. The caecum was chosen as the gut habitat for sampling because it is an anatomically distinct structure, located between the distal small intestine and colon, that is colonized with sufficient quantities of a readily harvested microbiota for metagenomic analysis. The predicted increased capacity for dietary energy harvest by the *ob/ob* microbiome was subsequently validated using biochemical assays and by transplantation of lean and obese caecal microbiotas into germ-free wild-type mouse recipients. These transplantation experiments illustrate the power of marrying metagenomics to gnotobiotics to discover how microbial communities encode traits that markedly affect host biology.

Shotgun sequencing of microbiomes

Bulk DNA was prepared from the caecal contents of two *ob/ob* and *+/+* littermate pairs. A lean *ob/+* mouse from one of the litters was also studied. All caecal microbial community DNA samples were analysed using a 3730xl capillary sequencer (10,500 ± 431 (s.e.m.) unidirectional reads per data set; 752 ± 13.8 (s.e.m.) nucleotides per read; 39.5 Mb from all five plasmid libraries). Material from one of the two obese and lean sibling pairs was also analysed using a highly parallel 454 Life Sciences GS20 pyrosequencer¹¹: three runs for the *+/+* mouse (known as lean1), and two runs for its *ob/ob* littermate (*ob1*) produced a total of 160 Mb of sequence (345,000 ± 23,500 (s.e.m.) unidirectional reads per run; 93.1 ± 1.56 (s.e.m.) nucleotides per read) (Supplementary Tables 1–3). Both sequencing platforms have unique advantages and limitations: capillary sequencing allows more confident gene calling (Supplementary Fig. 1) but is affected by cloning bias, whereas pyrosequencing can achieve higher sequence coverage with no cloning bias, but produces shorter reads

¹Center for Genome Sciences, and ²Genome Sequencing Center, Washington University, St. Louis, Missouri 63108, USA.

(Supplementary Table 2). The three pyrosequencer runs of the lean1 caecal microbiome (94.9 Mb) yielded $0.44\times$ coverage (on the basis of PROmer sequence alignments¹²) of the 3730xl-derived sequences obtained from the same sample (8.23 Mb), whereas the two pyrosequencer runs of the microbiome of its *ob/ob* littermate (*ob1*; 65.4 Mb) produced $0.32\times$ coverage of the corresponding 3730xl sequences (8.19 Mb).

Taxonomic analysis of microbiomes

Environmental gene tags (EGTs) are defined as sequencer reads assigned to the NCBI non-redundant, Clusters of Orthologous Groups¹³ (COG), or Kyoto Encyclopedia of Genes and Genomes¹⁴ (KEGG) databases (Fig. 1a; Supplementary Fig. 2; Supplementary Table 4). Averaging results from all data sets, 94% of the EGTs assigned to the non-redundant database were bacterial, 3.6% were eukaryotic (0.29% *Mus musculus*; 0.36% fungal), 1.5% were archaeal (1.4% Euryarchaeota; 0.07% Crenarchaeota), and 0.61% were viral (0.57% double stranded DNA viruses) (Supplementary Table 5). The relative abundance of the eight bacterial divisions identified from EGTs and 16S rRNA gene fragments was comparable to our previous PCR-derived, 16S-rRNA-gene-sequence-based surveys of

these caecal samples, including the increased ratio of Firmicutes to Bacteroidetes in obese versus lean littermates (Supplementary Fig. 2). In addition, comparisons of the lean1 and *ob1* reads obtained with the pyrosequencer against the finished genome of *Bacteroides thetaiotaomicron* ATCC29148¹, and a deep draft genome assembly of *Eubacterium rectale* ATCC33656 (50% of total contig bases present in contigs ≥ 75.9 kb; <http://gordonlab.wustl.edu/supplemental/Turnbaugh/obob/>) provided independent confirmation of the greater relative abundance of Firmicutes in the *ob/ob* microbiota. These organisms were selected for comparison because both are prominently represented in the normal human distal gut microbiota¹⁰ and species related to *B. thetaiotaomicron* (Bacteroidetes division) and *E. rectale* (Firmicutes division) are members of the normal mouse distal gut microbiota⁶. The ratio of sequences homologous to the *E. rectale* versus *B. thetaiotaomicron* genome was 7.3 in the *ob1* caecal microbiome compared with 1.5 in the lean1 microbiome.

Intriguingly, there were more EGTs that matched Archaea (Euryarchaeota and Crenarchaeota) in the caecal microbiome of *ob/ob* mice compared with their lean *ob/+* and *+/+* littermates (binomial test of pooled obese versus pooled lean capillary-sequencing-derived microbiomes, $P < 0.001$; Supplementary Table 5).

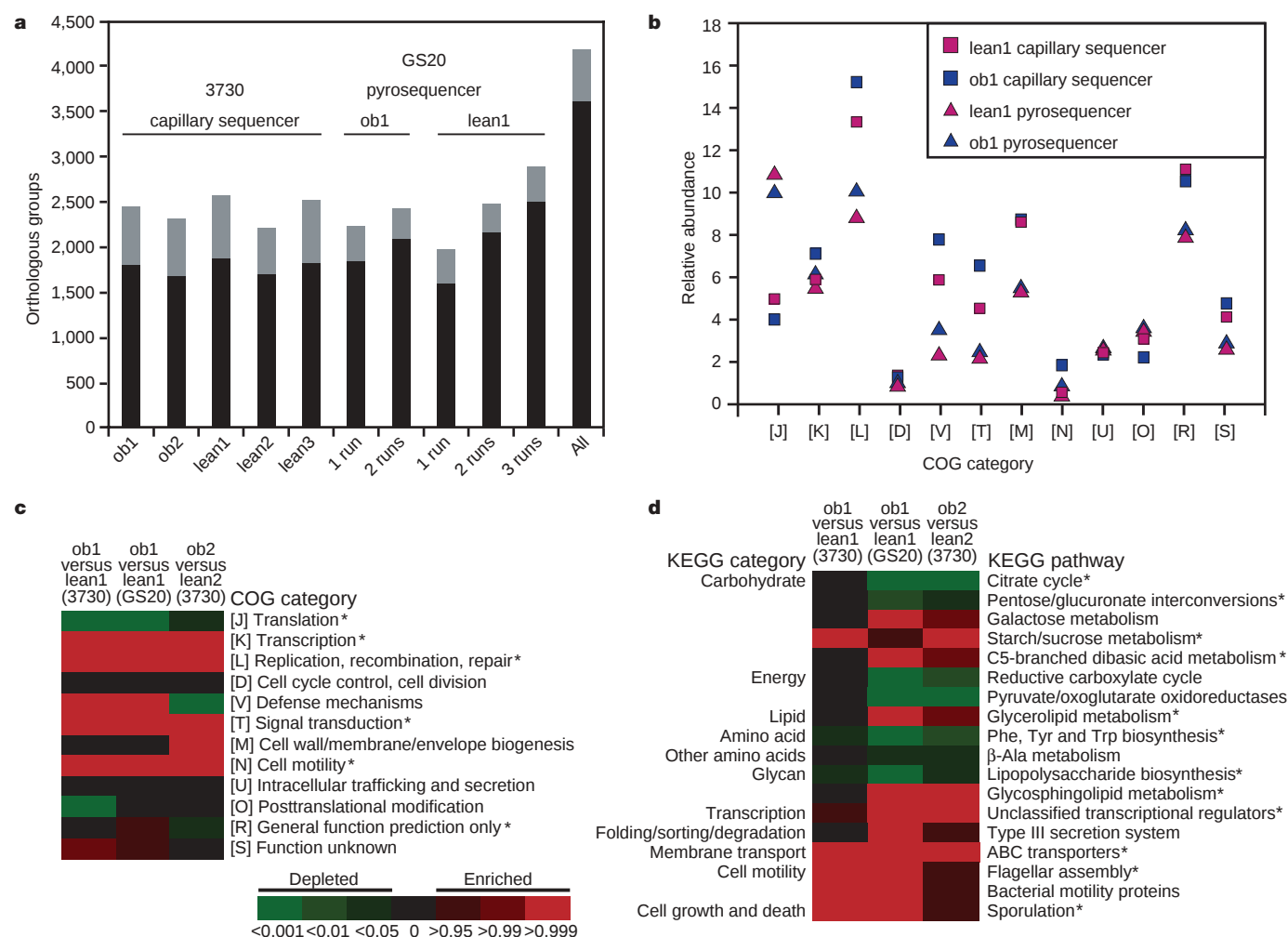


Figure 1 | Comparison of data sets obtained from the caecal microbiomes of obese and lean littermates. **a**, Number of observed orthologous groups in each caecal microbiome. Black indicates the number of observed groups. Grey indicates the number of predicted missed groups. **b**, Relative abundance of a subset of COG categories (BLASTX, e -value $< 10^{-5}$) in the lean1 (red) and *ob1* (blue) caecal microbiome, characterized by capillary- and pyro-sequencers (squares and triangles, respectively). **c**, **d**, A subset of COG categories (**c**) and all KEGG pathways (**d**) consistently enriched or depleted in the caecal microbiomes of both obese mice compared with their

lean littermates. Red denotes enrichment and green indicates depletion on the basis of a cumulative binomial test (brightness indicates the level of significance). Black indicates pathways whose representation is not significantly different. Asterisks indicate groups that were consistently enriched or depleted between both sibling pairs using a more stringent EGT assignment strategy (e -value $< 10^{-8}$). For additional details see Supplementary Discussion; Supplementary Figs 5 and 6, and Supplementary Tables 6, 8 and 9.

Methanogenic archaea increase the efficiency of bacterial fermentation by removing one of its end products, H_2 . Our recent studies of gnotobiotic normal mice colonized with the principal methanogenic archaeon in the human gut, *Methanobrevibacter smithii*, and/or *B. thetaiotaomicon* revealed that co-colonization not only increases the efficiency, but also changes the specificity of bacterial polysaccharide fermentation, leading to a significant increase in adiposity compared with mice colonized with either organism alone¹⁵.

Comparative metagenomic analysis

Using reciprocal TBLASTX comparisons, we found that the Firmicutes-enriched microbiomes from *ob/ob* hosts clustered together, as did lean microbiomes with low Firmicutes to Bacteroidetes ratios (Fig. 2a). Likewise, Principal Component Analysis of EGT assignments to KEGG pathways revealed a correlation between host genotype and the gene content of the microbiome (Fig. 2b).

Reads were then assigned to COGs and KOs (KEGG orthology terms) by BLASTX comparisons against the STRING-extended COG database¹³, and the KEGG Genes database¹⁴ (version 37). We tallied the number of EGTs assigned to each COG or KEGG category, and used the cumulative binomial distribution³, and a bootstrap analysis^{16,17}, to identify functional categories with statistically significant differences in their representation in both sets of obese and lean littermates. As noted above, capillary sequencing requires cloned DNA fragments; the pyrosequencer does not, but produces relatively short read lengths. These differences are a likely cause of the shift in relative abundance of several COG categories obtained using the two sequencing methods for the same sample (Fig. 1b). Nonetheless, comparisons of the caecal microbiomes of lean versus obese littermates sequenced with either method revealed similar differences in their functional profiles (Fig. 1c).

The *ob/ob* microbiome is enriched for EGTs encoding many enzymes involved in the initial steps in breaking down otherwise indigestible dietary polysaccharides, including KEGG pathways for starch/sucrose metabolism, galactose metabolism and butanoate metabolism (Fig. 1d; Supplementary Fig. 3 and Supplementary Table 6). EGTs representing these enzymes were grouped according to their functional classifications in the Carbohydrate Active Enzymes (CAZy) database (<http://afmb.cnrs-mrs.fr/CAZY/>). The *ob/ob* microbiome is enriched ($P < 0.05$) for eight glycoside hydrolase

families capable of degrading dietary polysaccharides including starch (CAZy families 2, 4, 27, 31, 35, 36, 42 and 68, which contain α -glucosidases, α -galactosidases and β -galactosidases). Finished genome sequences of prominent human gut Firmicutes have not been reported. However, our analysis of the draft genome of *E. rectale* has revealed 44 glycoside hydrolases, including a significant enrichment for glycoside hydrolases involved in the degradation of dietary starches (CAZy families 13 and 77, which contain α -amylases and amylomaltases; $P < 0.05$ on the basis of a binomial test of *E. rectale* versus the finished genomes of Bacteroidetes—*Bacteroides thetaiotaomicon* ATCC29148, *B. fragilis* NCTC9343, *B. vulgatus* ATCC8482 and *B. distasonis* ATCC8503).

EGTs encoding proteins that import the products of these glycoside hydrolases (ABC transporters), metabolize them (for example, α - and β -galactosidases KO7406/7 and KO1190, respectively), and generate the major end products of fermentation, butyrate and acetate (pyruvate formate-lyase, KO0656, and other enzymes in the KEGG 'Butanoate metabolism' pathway; and formate-tetrahydrofolate ligase, KO1938, the second enzyme in the homoacetogenesis pathway for converting CO_2 to acetate) are also significantly enriched in the *ob/ob* microbiome (binomial comparison of pyrosequencer-derived *ob1* and *lean1* data sets, $P < 0.05$) (Fig. 1d; Supplementary Fig. 3 and Supplementary Table 6).

As predicted from our comparative metagenomic analyses, the *ob/ob* caecum has an increased concentration of the major fermentation end-products butyrate and acetate (Fig. 3a). This observation is also consistent with the fact that many Firmicutes are butyrate producers^{18–20}. Moreover, bomb calorimetry revealed that *ob/ob* mice have significantly less energy remaining in their faeces relative to their lean littermates (Fig. 3b).

Microbiota transplantation

We performed microbiota transplantation experiments to test directly the notion that the *ob/ob* microbiota has an increased capacity to harvest energy from the diet and to determine whether increased adiposity is a transmissible trait. Adult germ-free C57BL/6J mice were colonized (by gavage) with a microbiota harvested

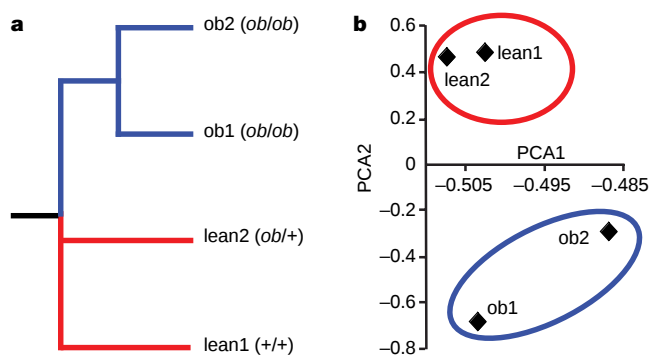


Figure 2 | Microbiomes cluster according to host genotype. **a**, Clustering of caecal microbiomes of obese and lean sibling pairs based on reciprocal TBLASTX comparisons. All possible reciprocal TBLASTX comparisons of microbiomes (defined by capillary sequencing) were performed from both lean and obese sibling pairs. A distance matrix was then created using the cumulative bitscore for each comparison and the cumulative score for each self-self comparison. Microbiomes were subsequently clustered using NEIGHBOUR (PHYLIP version 3.64). **b**, Principal Component Analysis (PCA) of KEGG pathway assignments. A matrix was constructed containing the number of EGTs assigned to each KEGG pathway in each microbiome (includes KEGG pathways with $>0.6\%$ relative abundance in at least two microbiomes, and a standard deviation >0.3 across all microbiomes), PCA was performed using Cluster3.0 (ref. 25), and the results graphed along the first two components.

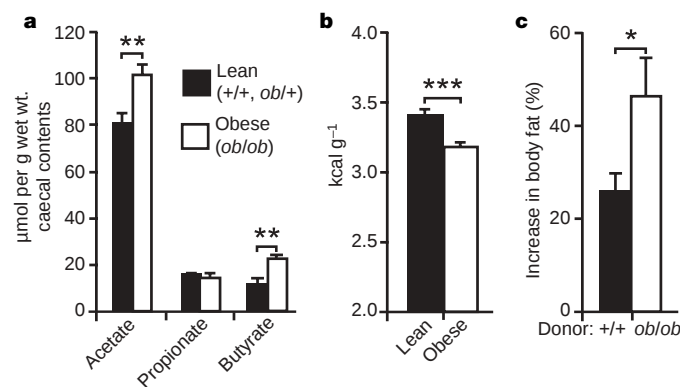


Figure 3 | Biochemical analysis and microbiota transplantation experiments confirm that the *ob/ob* microbiome has an increased capacity for dietary energy harvest. **a**, Gas-chromatography mass-spectrometry quantification of short-chain fatty acids in the caeca of lean ($n = 4$) and obese ($n = 5$) conventionally raised C57BL/6J mice. **b**, Bomb calorimetry of the faecal gross energy content (kcal g^{-1}) of lean ($+/+$, $ob/+$; $n = 9$) and obese (ob/ob ; $n = 13$) conventionally raised C57BL/6J mice. **c**, Colonization of germ-free wild-type C57BL/6J mice with a caecal microbiota harvested from obese donors (ob/ob ; $n = 9$ recipients) results in a significantly greater percentage increase in total body fat than colonization with a microbiota from lean donors ($+/+$; $n = 10$ recipients). Total body fat content was measured before and after a two-week colonization, using dual-energy X-ray absorptiometry. Mean values $\pm$ s.e.m. are plotted. Asterisks indicate significant differences (two-tailed Student's *t*-test of all datapoints, $*P < 0.05$, $**P \leq 0.01$, $***P < 0.001$).

from the caecum of obese (*ob/ob*) or lean (+/+) donors (1 donor and 4–5 germ-free recipients per treatment group per experiment; two independent experiments). 16S-rRNA-gene-sequence-based surveys confirmed that the *ob/ob* donor microbiota had a greater relative abundance of Firmicutes compared with the lean donor microbiota (Supplementary Fig. 4 and Supplementary Table 7). Furthermore, the *ob/ob* recipient microbiota had a significantly higher relative abundance of Firmicutes compared with the lean recipient microbiota ($P < 0.05$, two-tailed Student's *t*-test). UniFrac analysis²¹ of 16S rRNA gene sequences obtained from the recipients' caecal microbiotas revealed that they cluster according to the input donor community (Supplementary Fig. 4): that is, the initial colonizing community structure did not exhibit marked changes by the end of the two-week experiment. There was no statistically significant difference in (1) chow consumption over the 14-day period (55.4 ± 2.5 g (*ob/ob*) versus 54.0 ± 1.2 g (+/+); caloric density of chow, 3.7 kcal g⁻¹), (2) initial body fat (2.7 ± 0.2 g for both groups as measured by dual-energy X-ray absorptiometry), or (3) initial weight between the recipients of lean and obese microbiotas. Strikingly, mice colonized with an *ob/ob* microbiota exhibited a significantly greater percentage increase in body fat over two weeks than mice colonized with a +/+ microbiota (Fig. 3c; 47 ± 8.3 versus 27 ± 3.6 percentage increase or 1.3 ± 0.2 versus 0.86 ± 0.1 g fat (dual-energy X-ray absorptiometry): at 9.3 kcal g⁻¹ fat, this corresponds to a difference of 4 kcal or 2% of total calories consumed).

Discussion

The primary cause of obesity in the *ob/ob* mouse model is increased food consumption due to leptin deficiency. We have used this model to provide direct experimental evidence that at least one type of obesity-associated gut microbiome has an increased capacity for energy harvest from the diet. This finding provides support for the more general concept that the gut microbiome should be considered as a set of genetic factors that, together with host genotype and lifestyle (energy intake and expenditure), contribute to the pathophysiology of obesity. Yet to be answered are the questions of what mechanisms are responsible for mediating the linkage between the relative abundance of the Bacteroidetes to Firmicutes divisions and adiposity in both mice and humans, and to what extent is this relationship self-perpetuating?

Energy balance is an equilibrium between the amount of energy taken in as food and the amount expended during resting metabolism, as well as the thermic effect of food, physical activity, and loss in the faeces and urine. The alteration in efficiency of energy harvest from the diet produced by changes in gut microbial ecology does not have to be great to contribute to obesity, given that small changes in energy balance, over the course of a year, can result in significant changes in body weight²². We are aware of only one report showing that obese humans may have an increased capacity to absorb energy from their diet: in this case, analysis of four lean and four obese individuals given three different diets (high protein/high fat, 'average', and high carbohydrate) revealed that on average, the obese individuals lost less energy to stool compared with their lean counterparts. However, the differences did not achieve statistical significance²³.

Our study in mice demonstrates the feasibility and utility of applying comparative metagenomics to mouse models of human physiologic or pathophysiologic states in order to understand the complex interplay between host genetics, microbial community gene content and the biological properties of the resulting 'superorganism'. As such, it opens the door for comparable investigations of the interactions between humans and their microbial communities, including whether there is a core set of genes associated with the microbiomes of obese versus lean individuals; whether these genes are transmitted from mothers to their offspring; what genetic or behavioural traits of the host can reshape the community²⁴; how the microbiome changes as body mass index changes within an individual; the degree to which these changes correlate with energy harvest from their diets; and

whether germ-free mice can be used as a bioassay to compare the energy harvesting activities encoded in human gut microbiomes. Our results indicate that if the gut microbiome of obese humans is comparable to that of obese mice, then it may be a biomarker, a mediator and a new therapeutic target for people suffering from this increasingly worldwide disease.

METHODS

DNA was isolated from the caeca of *ob/ob*, *ob/+* and +/+ littermates using a bead beater to mechanically disrupt cells, followed by phenol–chloroform extraction. DNA was sequenced using 3730xl capillary- and GS20 pyro-sequencers: in the case of the latter, DNA was purified further using the Qiaquick gel extraction kit (Qiagen).

Individual reads in the resulting 199.8 Mb data set were directly compared with each other and to reference sequenced gut microbial genomes using MUMmer¹². Taxonomic assignments were made on the basis of BLASTX searches of the non-redundant database (*e*-value $< 10^{-5}$) and alignment of 16S gene fragments. Reads were also assigned to EGTs (environmental gene tags) by BLASTX searches against the non-redundant database, STRING-extended COG¹³, and KEGG¹⁴ (v37) databases. Microbiomes from each animal were clustered according to reciprocal TBLASTX comparisons and their EGT assignments to KEGG pathways. Statistically enriched or depleted COG and KEGG groups were identified using bootstrap^{16,17} and cumulative binomial³ analyses. For the binomial analysis, the probability of observing 'n₁' EGT assignments to a given group in microbiome 1, given 'N₁' EGT assignments to all groups in microbiome 1, was calculated using the cumulative binomial distribution and an expected probability equal to 'n₂/N₂' (the number of EGTs assigned to a given group in microbiome 2 divided by the total number of EGTs assigned to all groups in microbiome 2). Detailed descriptions of these methods and techniques for (1) measuring short-chain fatty acids in caecal samples by gas-chromatography mass-spectrometry, (2) bomb calorimetry of faecal samples, (3) transplanting the caecal microbiota of C57BL/6J *ob/ob* or +/+ donors into 8–9-week-old germ-free +/+ C57BL/6J recipients, (4) measuring the total body fat of transplant recipients, before and after colonization, by dual-energy X-ray absorptiometry, and (5) performing 16S-rRNA-gene-sequence-based surveys of the input (donor) and output (recipient) caecal microbiotas are provided in Supplementary Information.

Received 8 October; accepted 7 November 2006.

- Xu, J. *et al.* A genomic view of the human–*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**, 2074–2076 (2003).
- Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host–bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
- Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006).
- Sonnenburg, J. L. *et al.* Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005).
- Backhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
- Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
- Ley, R. E. *et al.* Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl. Environ. Microbiol.* **72**, 3685–3695 (2006).
- Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837–848 (2006).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity. *Nature* doi:10.1038/nature4441023a (this issue).
- Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
- Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
- Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
- von Mering, C. *et al.* STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* **33**, D433–D437 (2005).
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).
- Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc. Natl Acad. Sci. USA* **103**, 10011–10016 (2006).
- DeLong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496–503 (2006).
- Rodríguez-Brito, B., Rohwer, F. & Edwards, R. An application of statistics to comparative metagenomics. *BMC Bioinformatics* **7**, 162 (2006).
- Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 1615–1620 (2002).

19. Barcenilla, A. *et al.* Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* **66**, 1654–1661 (2000).
20. Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS Microbiol. Lett.* **217**, 133–139 (2002).
21. Lozupone, C., Hamady, M. & Knight, R. UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**, 371 (2006).
22. Flegal, K. M. & Troiano, R. P. Changes in the distribution of body mass index of adults and children in the US population. *Int. J. Obes. Relat. Metab. Disord.* **24**, 807–818 (2000).
23. Webb, P. & Annis, J. F. Adaptation to overeating in lean and overweight men and women. *Hum. Nutr. Clin. Nutr.* **37**, 117–131 (1983).
24. Rawls, J. F., Mahowald, M. A., Ley, R. E. & Gordon, J. I. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* **127**, 423–433 (2006).
25. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank our colleagues B. Samuel, F. Backhed, D. O'Donnell, M. Karlsson, M. Hickenbotham, K. Haub, L. Fulton, J. Crowley, T. Coleman, C. Semenkovich, V. Markowitz and E. Szeto for their assistance. This work was supported by grants from the NIH and the W.M. Keck Foundation.

Author Information This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the project accession AATA00000000–AATF00000000. The version described in this paper is the first version, AATA01000000–AATF01000000. All 454 GS20 reads have been deposited in the NCBI Trace Archive. PCR-derived 16S rRNA gene sequences are deposited in GenBank under the accession numbers EF95962–100118. Annotated sequences are also available for further analysis in IMG/M (<http://img.jgi.doe.gov/m>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.I.G. (jgordon@wustl.edu).

ARTICLES

Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis

Irene Noguera-Troise¹, Christopher Daly¹, Nicholas J. Papadopoulos¹, Sandra Coetzee¹, Pat Boland¹, Nicholas W. Gale¹, Hsin Chieh Lin¹, George D. Yancopoulos¹ & Gavin Thurston¹

Tumour growth requires accompanying expansion of the host vasculature, with tumour progression often correlated with vascular density. Vascular endothelial growth factor (VEGF) is the best-characterized inducer of tumour angiogenesis. We report that VEGF dynamically regulates tumour endothelial expression of Delta-like ligand 4 (Dll4), which was previously shown to be absolutely required for normal embryonic vascular development. To define Dll4 function in tumour angiogenesis, we manipulated this pathway in murine tumour models using several approaches. Here we show that blockade resulted in markedly increased tumour vascularity, associated with enhanced angiogenic sprouting and branching. Paradoxically, this increased vascularity was non-productive—as shown by poor perfusion and increased hypoxia, and most importantly, by decreased tumour growth—even for tumours resistant to anti-VEGF therapy. Thus, VEGF-induced Dll4 acts as a negative regulator of tumour angiogenesis; its blockade results in a striking uncoupling of tumour growth from vessel density, presenting a novel therapeutic approach even for tumours resistant to anti-VEGF therapies.

Tumour growth depends on expansion of the host vasculature into the tumour, through the process of tumour angiogenesis¹. The connection between tumour growth and angiogenesis prompted the development of several approaches to limit tumour angiogenesis and thus control tumour growth. The best validated of these approaches involves blockade of the VEGF pathway. Blockade of VEGF controls tumour growth in numerous preclinical models^{2,3}, and recent results show that potent blockers of VEGF can completely prevent tumour angiogenesis in some models, thereby severely inhibiting tumour growth⁴. The promise of VEGF-blocking approaches has recently been realized in the clinic, as a VEGF-blocking antibody has been shown to have important effects on tumour progression and overall survival in cancer patients^{5,6}. However, despite the critical role for VEGF in tumour angiogenesis, it is also clear that in some cases tumour growth and angiogenesis can proceed even in the face of potent VEGF blockade^{7–9}. Thus, additional angiogenesis-targeted therapies are necessary for tumours resistant to VEGF blockade.

Mouse genetic studies have demonstrated that, in addition to the VEGF pathway, other signalling pathways are also required for normal embryonic vascular development (for reviews, see refs 10–12), raising the possibility that these pathways may also be important during tumour angiogenesis. One signalling pathway implicated in vascular development by gene deletion studies is the Notch pathway¹³. On binding a transmembrane ligand from the Delta/Jagged families, Notch transmembrane receptors generally provide signals to guide cell fate decisions^{14,15}. In particular, Delta-like ligand 4 (Dll4) is absolutely required for normal vascular development^{16–18} and is strongly expressed in tumour vessels^{17,19,20}.

To determine whether the Dll4/Notch pathway has a role during tumour angiogenesis, we manipulated this pathway in experimental tumour models in mice using a variety of genetic and pharmacologic approaches. We report that the Dll4/Notch pathway is a critical negative regulator of tumour angiogenesis, acting to restrain excessive VEGF-induced vascular sprouting and angiogenesis. Increased Dll4/Notch activity resulted in decreased tumour vascular density, whereas blockade of activity resulted in markedly increased vessel

density. Paradoxically, this increased vascularity seemed to be non-productive and resulted in decreased tumour growth, even for tumours that are resistant to anti-VEGF therapy. Our findings provide a striking example of an uncoupling of tumour growth from tumour vascular density, and support the model that the Dll4/Notch pathway normally acts as a negative regulator of angiogenic sprouting induced by VEGF or other pathways.

RESULTS

VEGF induces high expression of Dll4 in tumour vessels. To confirm and extend earlier reports that Dll4 expression was markedly and specifically induced in blood vessels during tumour angiogenesis^{17,19,20}, we used a combination of Dll4 detection approaches in two different tumour models. First, we exploited ‘*Dll4* reporter mice’ in which a β -galactosidase reporter gene was driven by the *Dll4* promoter¹⁷. Lewis lung carcinomas implanted into these mice showed strong reporter-based staining of tumour vessels, apparently at higher levels than of vessels in surrounding normal tissue (Fig. 1a, b). At higher resolution, immunostaining for the β -galactosidase reporter protein and comparison with adjacent sections in which all vessels were immunostained with CD-31/PECAM antibodies revealed strong *Dll4* reporter expression in tumour blood vessels and relatively weak staining in adjacent subcutaneous and dermal blood vessels¹⁷ (Supplementary Fig. 1). To confirm that the β -galactosidase/*Dll4* reporter (*Dll4*–*LacZ*) construct marked sites of Dll4 protein, we generated polyclonal antibodies to murine Dll4; these antibodies also immunostained tumour blood vessels selectively (Fig. 1c, d). These findings were further confirmed by *in situ* hybridization for *Dll4* messenger RNA, which revealed prominent expression in the vessels of C6 glioma tumours¹⁹ (Supplementary Fig. 1). Thus, Dll4 is indeed specifically expressed in the tumour vasculature, particularly in the smaller vessels. Moreover, expression of Dll4 was dependent on continuous VEGF signalling, because blockade of VEGF with VEGF Trap, a recombinant soluble receptor that potently blocks VEGF-A and placental growth factor (PlGF)⁴, caused a rapid and marked decrease in the expression of Dll4 by tumour vessels (Fig. 1e).

¹Regeneron Research Laboratories, 777 Old Saw Mill River Road, Tarrytown, New York 10591, USA.

Activating and blocking the Dll4/Notch pathway in tumours. To manipulate the Dll4/Notch pathway in tumours, we first exploited a retroviral approach to overexpress forms of Dll4, which we reasoned would serve as blockers or activators, in tumour cells. On the basis of previous studies that used soluble versions of Dll to inhibit Notch signalling²¹, we generated retroviral vectors encoding a soluble dimerized version of Dll4 in which the extracellular region of Dll4 was fused to the human IgG1 Fc constant region (termed Dll4-Fc) as a presumed blocker, as well as full-length membrane-bound Dll4 that presumably would act as an activator; these constructs were transduced into rat C6 tumour cells to produce C6 Dll4-Fc and C6 Dll4 cells. Importantly, C6 tumour cells that overexpressed Dll4-Fc or Dll4 did not have different growth characteristics *in vitro* than control tumour cells (data not shown).

The expected changes on Notch signalling in the host (mouse) stroma of subcutaneously implanted C6 rat tumour cells were confirmed by quantitative messenger RNA analyses using probes specific for the mouse versions (to detect expression in the host stromal cells and not in the rat tumour cells) of three genes that are characteristic targets of Notch signalling (*HES1*, *HEY2* and *NRARP*)^{22–27}. That is, in C6 Dll4-Fc tumours, host Notch signalling was consistently reduced as reflected by decreased levels of these target genes, whereas in

C6 Dll4 tumours the Notch pathway was activated (Supplementary Fig. 3). Similarly, in co-culture studies in which the transduced rat tumour cells were mixed with human umbilical vein endothelial cells, expression levels of the *HES1*, *HEY2* and *NRARP* genes in the cultured endothelial cells (assayed using human specific probes so as to specifically detect the endothelial versions of these transcripts) were reduced by co-culture with C6 Dll4-Fc cells and induced by C6 Dll4 cells (Supplementary Fig. 4). Finally, treatment of cultured human umbilical vein endothelial cells with purified Dll4-Fc protein rapidly and consistently repressed Notch signalling, as reflected by decreased expression of these target genes (Supplementary Fig. 5).

Blockade of Dll4/Notch results in decreased tumour growth. To explore the effects of Dll4/Notch pathway manipulation in tumours, we next examined tumours derived from C6 Dll4-Fc and C6 Dll4 cells for their vascular morphologies and tumour growth rates. Strikingly, Dll4-Fc and full-length Dll4 seemed to promote reciprocal changes in the tumour vasculature: the vasculature in C6 Dll4-Fc tumours was much more highly branched and had more fine interconnections than that of control tumours (Fig. 2a–f). The leading front of the vessels was replete with sprouts and filopodia (Fig. 2e, h). In contrast, the vasculature in C6 Dll4 tumours was notably

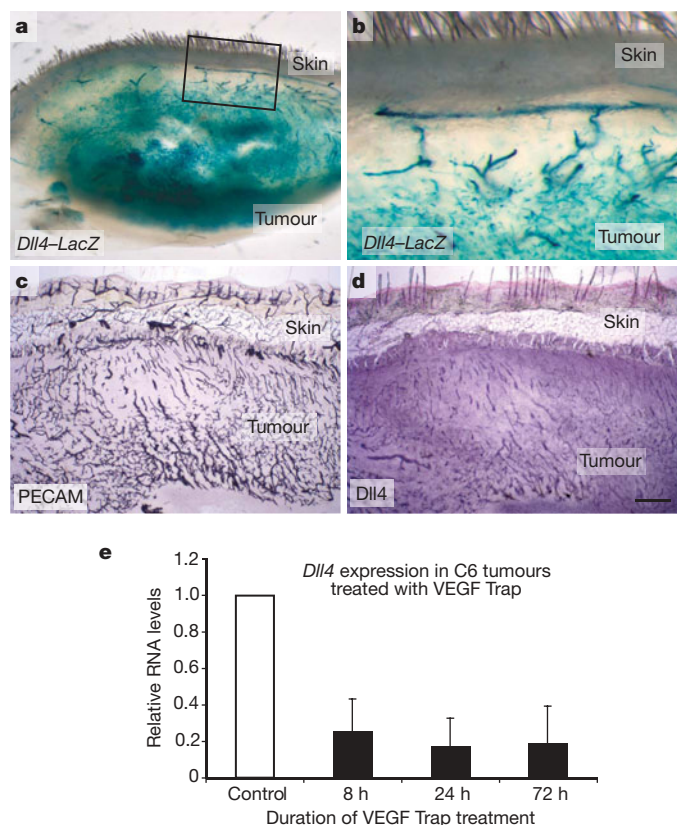


Figure 1 | Dll4 is expressed in tumour vessels, and its expression is dependent on VEGF signalling. **a, b**, Lewis lung tumours were grown subcutaneously in heterozygous *Dll4*-targeted mice in which the *Dll4* gene was replaced with the gene encoding β -galactosidase. **a**, *Dll4* expression, revealed by staining for LacZ (blue), was strong in tumour tissue, but weak in normal skin tissue. **b**, Higher magnification image showing *Dll4* expression in tumour blood vessels. **c**, Section of tumour showing immunoreactivity for CD31/PECAM is equally strong in tumour vessels and adjacent skin vessels. **d**, In contrast, immunoreactivity for Dll4 is strong in tumour vessels and weaker in vessels of the adjacent skin. Scale bar: 400 μ m (**c, d**). **e**, Expression of *Dll4* in tumour vessels is decreased by VEGF blockade. Rat C6 tumours grown subcutaneously were treated for the indicated times with VEGF Trap. Tumours were analysed for expression of murine *Dll4* by microarray analysis. Data show mean $\pm$ s.d. of *Dll4* expression compared with control C6 tumours from three tumours per time point ($n = 3$).

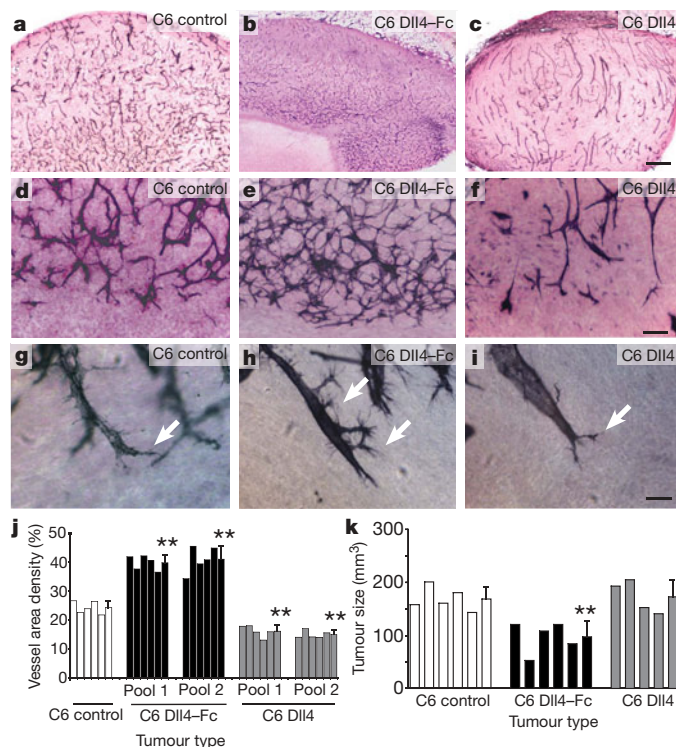


Figure 2 | Blockade of Dll4/Notch signalling results in smaller C6 tumours with increased vessel density. **a–i**, Micrographs showing tumours stained for CD31 (black) in control C6 tumours (**a, d, g**), in C6 tumours overexpressing full-length Dll4 (**c, f, i**). Micrographs show tumour sections at low, medium and high magnification. C6 Dll4-Fc tumours contain a dense network of vessel structures, with numerous cellular processes, particularly at the leading front (**b, e**). Reciprocally, C6 Dll4 tumours contain relatively sparse and unbranched vessels (**c, f**). The leading cells in the actively growing vascular front of Dll4-Fc tumours have more cellular processes than those in control tumours, and reciprocally, such cells in Dll4 tumours have fewer processes (arrows, **g–i**). Scale bars, 400 μ m (**a–c**); 100 μ m (**d–f**); 20 μ m (**g–i**). **j**, Quantification of vessel area density by morphometry shows increased vessel density in C6 Dll4-Fc tumours and decreased vessel density in C6 Dll4 tumours. **k**, C6 Dll4-Fc tumours are smaller than C6 GFP control tumours, whereas C6 Dll4 tumours are the same size. Quantification was done on two independently isolated pools of C6 Dll4-Fc and C6 Dll4 cells (results from a single pool are shown). Data from individual tumours are shown, as well as mean $\pm$ s.d. (**j, k**) for each group ($n = 4–5$); $^{**}P < 0.01$.

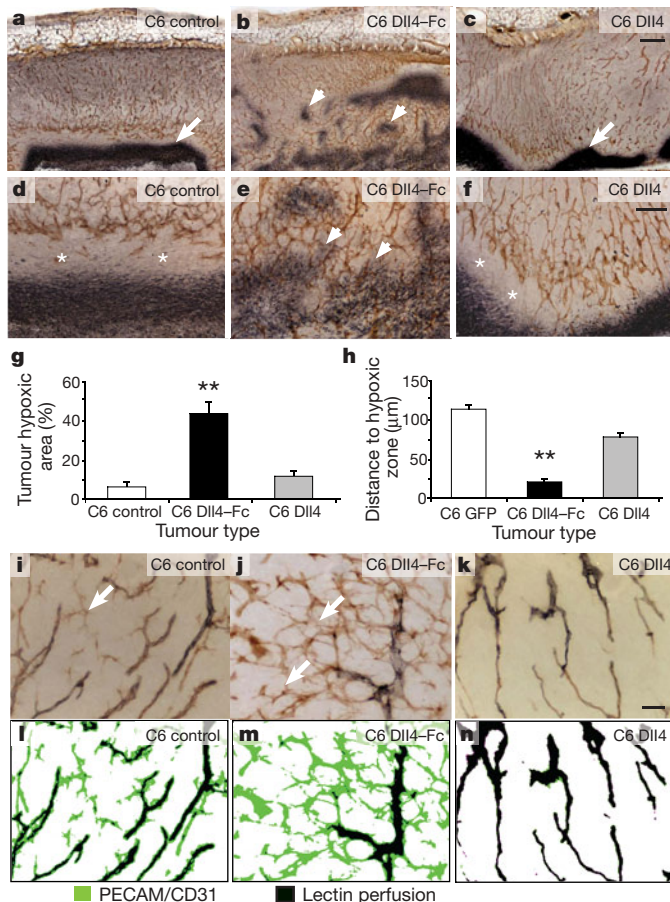


Figure 3 | Despite an increase in blood vessel density, tumours overexpressing Dll4-Fc have increased hypoxia and poor vascular perfusion. **a–f**, C6 tumours were stained for CD31/PECAM (brown) and for hypoxia (HypoxyProbe, black). Control C6 tumours of this size had no evident necrosis, and had very little hypoxia in the region of vascularized tumour, but did have a prominent rim of hypoxia at the base of the tumour (**a**, black staining, arrow) that was separated from the leading front of tumour vessels (asterisks in **d**). In comparison, C6 Dll4-Fc tumours contained increased areas of hypoxia in the region of vascularized tumour (arrowheads, **b**). In addition, hypoxia was found immediately adjacent to, or intermingled with, the leading front of tumour vessels (arrowheads, **e**). C6 Dll4 tumours had few areas of hypoxia in the region of vascularized tumour, and the hypoxic region (arrow, **c**) was typically separated from the leading front of vessels (asterisks, **f**), although more variability was seen than in control tumours. **g**, Quantification of hypoxic region within the vascularized tumour. C6 Dll4-Fc tumours contained significantly more hypoxia. **h**, Quantification of the distance between the leading front of vessels and the hypoxic zone, as marked with white asterisks in panels **d**, **e**, **f**. The distance to the hypoxic region was significantly less in C6 Dll4-Fc tumours than in control tumours. Values are mean $\pm$ s.d. (**g**, **h**) for each group ($n = 4–5$); ** $P < 0.01$. **i–n**, Comparison of vessel perfusion and overall CD31/PECAM immunoreactivity reveals decreased vessel function in C6 Dll4-Fc tumours. Tumour vessels were stained by *in vivo* intravascular injection of biotinylated lectin (stained in black) to mark perfused vessels, and by immersion in CD31/PECAM antibodies (stained in brown) to mark all vascular structures. **i–n**, Computer-generated colour images represent thresholds to show perfused vessels (black) and non-perfused vascular structures (green). Vascular structures in control C6 tumours (**i**, **l**) had a mixture of only brown staining, showing lack of perfusion (particularly in the fine processes, arrow) as well as a combination of black and brown staining showing perfused vessels. In C6 Dll4-Fc tumours (**j**, **m**), many vascular structures, particularly the numerous fine processes, were not perfused by intravascular tracer and thus stained only for CD31 (arrows). Reciprocally, in C6 Dll4 tumours (**k**, **n**), the unbranched vascular structures, which lack fine processes, were stained by both perfused lectin (black) and CD31 (brown), showing that vascular structures are functional. Scale bars, 100 μ m (**a–c**); 40 μ m (**d–f**); 20 μ m (**i–n**).

straighter and relatively unbranched (Fig. 2c, f), and relatively devoid of sprouts and filopodia (Fig. 2f, i).

These obvious morphologic changes were reflected by quantitation of vascular area densities in these tumours, with the C6 Dll4-Fc tumours exhibiting increased vascular density as compared with control tumours, whereas the C6 Dll4 tumours exhibited slightly decreased vascular density (Fig. 2j). Paradoxically, the effects of Dll4-Fc and full-length Dll4 on vascular density were opposite to those that might be expected with respect to their effects on tumour growth. Despite increased vascular density, C6 Dll4-Fc tumours were consistently smaller than control tumours, whereas C6 Dll4 tumours were not substantially different in size from control tumours for small tumours (Fig. 2k) or when C6 Dll4 tumours were harvested at larger sizes (Supplementary Fig. 6). These results have two important implications: first, that the Dll4/Notch pathway normally serves as a negative regulator of sprouting and branching activity during tumour angiogenesis, so that blockade of this pathway results in increased tumour angiogenesis; and second, that the increased angiogenesis resulting from blockade of this pathway is in some sense ‘non-productive’, such that it does not support more robust tumour growth and instead seems to be associated with reduced tumour growth.

Blockade of Dll4/Notch increases tumour hypoxia. To account for the apparent paradox of increased tumour vessel density and decreased tumour growth in C6 Dll4-Fc tumours, we examined the possibility that the increased network of vessels might not be

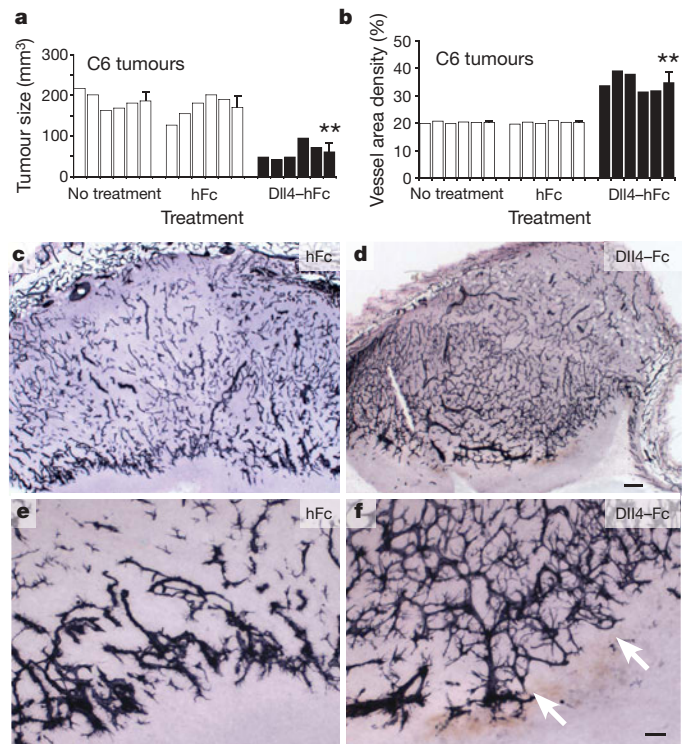


Figure 4 | Systemic delivery of Dll4-Fc using adenovirus results in smaller C6 tumours and increased vessel density, similar to effects of local tumour overexpression. **a**, C6 tumours were smaller in mice treated with systemically delivered Dll4-Fc, but not by control hFc. **b**, The vessel area density is increased in C6 tumours treated by systemic Dll4-Fc compared with untreated tumours or those treated with systemic hFc. The graphs in **a** and **b** show data from individual tumours, as well as mean $\pm$ s.d. for each group ($n = 5$). ** denotes significantly different than the control tumour group ($P < 0.01$). **c–f**, Micrographs showing tumour vessel morphology in control tumours (hFc; **c**, **e**) and in tumours treated systemically with Dll4-Fc (Dll4-Fc; **d**, **f**). The vessel density is increased in C6 Dll4-Fc tumours, particularly at the leading front (**f**, arrows). Tumour sections stained for CD31 (PECAM, black). Scale bars, 200 μ m (**c**, **d**); 50 μ m (**e**, **f**).

optimally functional. In agreement with this possibility, histologic assessment showed more extensive tumour hypoxia in C6 Dll4-Fc tumours than in control tumours (Fig. 3a, b; hypoxic regions stained in black). Moreover, whereas the hypoxic region in control tumours was separated from the growing front of tumour vessels by an avascular zone that was itself not hypoxic (corresponding to the oxygen diffusion distance; white asterisks, Fig. 3d), areas of hypoxia were interspersed with the tumour vasculature in C6 Dll4-Fc tumours (arrowheads, Fig. 3b, e), indicating that this vasculature was not efficiently delivering oxygen to the surrounding tumour. Quantitative analysis showed that Dll4-Fc tumours contained seven-fold more total hypoxic area within the vascularized tumour (Fig. 3g), and that the hypoxic rim was less separated from the leading front of tumour vessels (Fig. 3h). Corresponding values in the C6 Dll4 tumours were not substantially different from controls

(Fig. 3c), although there was more variability in the distance from the vascular front to the hypoxic rim (white asterisks, Fig. 3f).

The increase in tumour hypoxia in the C6 Dll4-Fc tumours suggested that the dense network of vessels was not fully perfused. We compared the distribution of vessel perfusion (marked by intravascular lectin as a tracer) with the immunohistochemical staining of endothelial cells (stained with CD31/PECAM-1 antibodies) as a marker of total vasculature. Most of the larger vessels of control tumours were perfused, although—as expected—these vessels were associated with some non-perfused sprouts and smaller vessels emanating from the larger vessels (Fig. 3i, brown reveals total vasculature, black reveals perfused vessels; Fig. 3l, computer-generated colour depiction, with green showing total vasculature and black showing perfused vessels). In contrast, many of the vascular processes in the C6 Dll4-Fc tumours were not perfused (Fig. 3j, m), suggesting that the increased vessel density seen in these tumours is not part of a functional vascular network. Reciprocally, the relatively straight and unbranched vessels seen in the C6 Dll4 tumours were almost completely perfused (Fig. 3k, n).

Together, the tumour hypoxia and perfusion analyses support the notion that the increased vascular network that results from inhibition of the Dll4/Notch pathway (by Dll4-Fc) is not optimally functional and is instead 'non-productive'.

Systemic Dll4-Fc decreases tumour growth. To confirm and extend the finding that locally produced Dll4-Fc promotes excessive angiogenesis that paradoxically blunts tumour growth, we used an adenoviral delivery approach to determine whether systemic Dll4-Fc could also produce this effect. Adenoviruses expressing Dll4-Fc, or human Fc (hFc) as control, were injected intravenously into mice at the time of implanting subcutaneous C6 tumours. The intravenously injected adenovirus infects hepatocytes, and in the case of Dll4-Fc or hFc, produced high serum levels of the encoded protein of $74 \pm 20 \mu\text{g ml}^{-1}$ (range of $50\text{--}100 \mu\text{g ml}^{-1}$, $n = 5$ mice). Circulating Dll4-Fc resulted in an approximately 70% reduction in the size of subcutaneous C6 tumours (Fig. 4a), whereas circulating control hFc had no effect. When examined by histology, circulating Dll4-Fc also caused an increase in the density of the tumour vessels (Fig. 4b). The overall increased vessel density produced by circulating Dll4-Fc was associated with a dense mesh of highly branched and sprouted vessels, particularly at the leading front (Fig. 4c–f), similar to that produced by Dll4-Fc overexpression in the tumour cells. Gene expression analysis confirmed that systemic Dll4-Fc suppressed the Notch pathway in the tumour vessel, as indicated by decreased expression of *HES1* and *NRARP* (data not shown). Thus, inhibition of Dll4/Notch signalling by either local or systemic Dll4-Fc results in smaller C6 tumours and excessive but apparently non-productive tumour angiogenesis. Importantly, the systemic treatment did not seem to have untoward effects on the host animals; in addition, preliminary analysis of normal tissues did not reveal obvious changes in tissue vascularity (Supplementary Information).

Dll4-Fc or Dll4-blocking antibody act in multiple tumour models.

To further extend the above findings, we used systemic injection of purified recombinant Dll4-Fc protein as the treatment, and tried additional tumour models. The above studies were carried out using C6 gliomas, which are relatively sensitive to the effects of VEGF blockade⁴, so we assessed the response to Dll4/Notch blockade in other tumours that are more resistant to VEGF blockade. In previous experiments using both bevacizumab (Avastin) and VEGF Trap, we had developed a model of HT1080 tumours (HT1080-(resistance model)RM), which is relatively resistant to both Avastin and VEGF Trap (G.T., I.N.-T. and J. Rudge, unpublished results). In contrast to Avastin and VEGF Trap, Dll4-Fc protein was quite effective in reducing the growth of HT1080-RM tumours (Fig. 5a). In a separate experiment (Fig. 5b), we assessed the growth curves of HT1080-RM tumours treated with Dll4-Fc, VEGF Trap or control protein (all at 25 mg kg^{-1} , three times per week; treatment began when tumours were approximately 100 mm^3 in size, arrow). Dll4-Fc

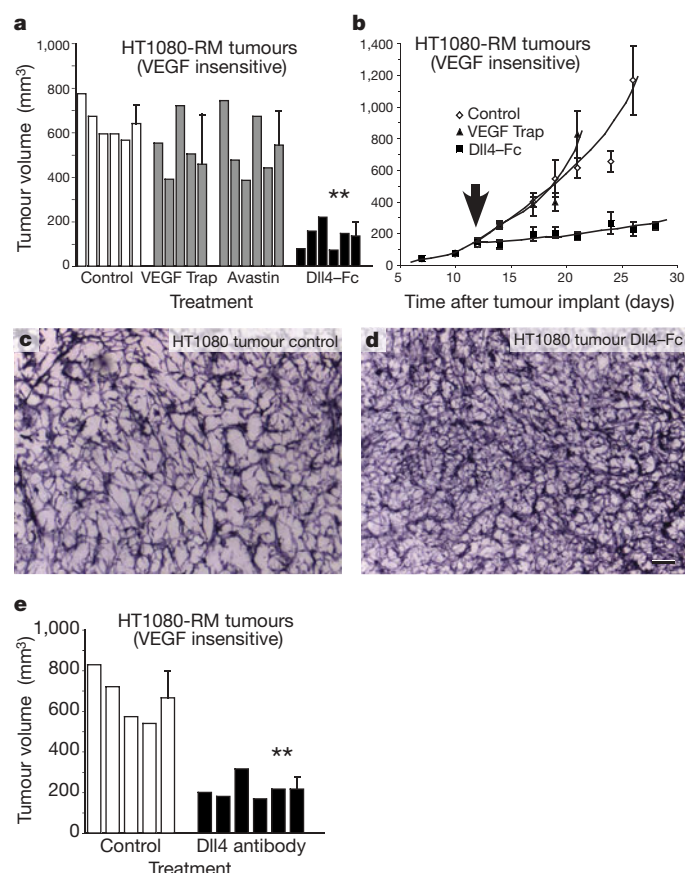


Figure 5 | Systemic delivery of Dll4-Fc or blocking Dll4 antibodies to mice bearing tumours that are resistant to blockade of VEGF results in decreased tumour growth and dramatic changes in tumour vessels. **a**, Size of HT1080-RM tumours treated with Dll4-Fc protein or VEGF Trap or bevacizumab (Avastin, all at 25 mg kg^{-1}). Tumours were treated with subcutaneous injections starting once tumours had reached approximately 100 mm^3 . Dll4-Fc treated tumours were smaller than controls, whereas those treated with blockers of VEGF were almost the same size as controls. **b**, Tumour growth curves from a separate experiment in which HT1080-RM tumours were treated from 100 mm^3 (arrow) with VEGF Trap, Dll4-Fc, or hFc (25 mg kg^{-1} , three times per week). Error bars are s.d. Whereas tumours treated with VEGF Trap grew as rapidly as control tumours, those treated with Dll4-Fc were restrained for at least 2 weeks of treatment. **c**, **d**, Dll4-Fc caused an increase in tumour vessel density and dramatic changes in vessel morphology in HT1080-RM tumours. Tumour sections were stained for CD31. Scale bar, $50 \mu\text{m}$ (**c**, **d**). **e**, Antibodies to Dll4 strongly suppressed tumour growth, similar to the effect of Dll4-Fc. HT1080-RM tumours were treated from a size of 100 mm^3 with polyclonal antibodies to Dll4 (10 mg kg^{-1} , three times per week) or with rabbit IgG as a control. The bar graphs show data from individual tumours, as well as the mean $\pm$ s.d. for each group ($n = 4\text{--}5$).

treatment resulted in a prolonged suppression of tumour growth, whereas VEGF Trap had almost no impact on tumour growth in this resistant tumour model. Dll4–Fc treatment of HT1080-RM tumours also had a marked effect on tumour vessels. The already-dense vasculature of control HT1080 tumours (Fig. 5c) was further increased by treatment with Dll4–Fc (Fig. 5d), inducing an apparent increase in vascular sprouting and branching, and an apparent disorganization to the network.

To verify the specificity of blocking with Dll4–Fc, we also generated polyclonal antibodies to the extracellular portion of Dll4, which inhibited the binding of Dll4 to Notch1 *in vitro* assays (Supplementary Fig. 7). Systemic treatment of mice bearing resistant HT1080-RM tumours with blocking Dll4 antibodies also reduced tumour growth (Fig. 5e) and caused marked changes in tumour vessels (data not shown).

As for HT1080 tumours, growth of mouse mammary tumours, which are also resistant to VEGF blockade (not shown), was strongly suppressed by systemic treatment with Dll4–Fc (Supplementary Fig. 8). Again, the vascularity of mouse mammary tumours was further increased by treatment with Dll4–Fc (Supplementary Fig. 8). Thus, as with C6 tumours, treatment of other tumours with systemic Dll4–Fc resulted in decreased tumour growth, accompanied by a denser and more highly branched tumour vasculature.

Discussion

To determine whether the Dll4/Notch pathway has a role during tumour angiogenesis, we manipulated this pathway in tumours using a variety of approaches. Our findings suggest that tumour-derived VEGF induces Dll4 expression in angiogenic endothelial cells as a critical negative regulator of vascular growth, acting to restrain excessive vascular sprouting and branching, and allowing angiogenesis to proceed at a productive rate (Supplementary Fig. 9). Thus, increasing Dll4/Notch activity resulted in decreased vascular density associated with less sprouting and branching of the vascular network. In contrast, Dll4/Notch blockade was associated with enhanced angiogenic sprouting and branching, resulting in a marked increase in tumour vessel density but a decrease in vessel function (Supplementary Fig. 9). Previously, angiogenesis-based treatment of tumours has focused on trying to block angiogenesis; however, our results using Dll4 blockade suggest an alternative approach based on promoting 'non-functionality' in the growing tumour vasculature. Although our studies suggest that VEGF blockade may be equally or more effective than Dll4 blockade in many tumour models, certain models that are resistant to VEGF blockade can still be sensitive to Dll4 blockers.

We, and others, have shown that Dll4 is specifically expressed in remodelling vessels and is the major Notch ligand in the vasculature. Thus, rather than a general blockade of the Notch pathway, specific blockade of Dll4 may lead to more specific disruption of tumour growth without significant impairment of Notch function in normal host tissues, and thus might be well tolerated in long-term treatments. It seems likely that biological therapeutic agents, which can be specific to a particular ligand or receptor in this complex pathway, may prove more potent and specific than more general pathway blockers, such as the γ -secretase inhibitors that not only block all Notch signalling but also other important γ -secretase-mediated signalling as well.

Our findings provide a striking example of an uncoupling of tumour growth from tumour vascular density. Although a large literature supports the notion that tumour growth rate may correlate with tumour vascular density, other studies argue that tumour angiogenesis must be regulated to be productive. For example, a recent study suggests that tumours may have higher vascular densities than is necessary to support their growth, and thus tumour angiogenesis may often exceed an optimally productive rate²⁸. Consistent with the concept of excessive tumour vessel density, some recent studies suggest that pruning of the vasculature might actually improve tumour

perfusion and oxygenation^{29–31}. The present studies with blockers of Dll4/Notch seem to provide the other side of the argument; in particular, that Dll4 blockade may further compromise tumour vasculature function by causing excessive non-productive angiogenesis, which can in turn inhibit tumour growth. The overall message seems to be that even tumour vascular networks require a regulated balance of growth factors to generate a hierarchy of well-organized and well-functioning vessels. VEGF clearly has a key angiogenic role in a wide variety of tumours, but Dll4 blockade may present a new therapeutic opportunity in cancer, and one that might be beneficial for patients with tumours that are resistant to anti-VEGF therapies.

METHODS

Engineered retroviruses. Retroviruses engineered to express green fluorescent protein (GFP) (control), GFP plus Dll4–hFc or GFP plus Dll4 were used to transduce C6 rat glioma tumour cells. Tumour cell pools, sorted by flow cytometry, were implanted subcutaneously in severe combined immunodeficient (SCID) mice (8–10 weeks old). Tumours were harvested and processed for histology and/or gene expression analysis.

Engineered adenoviruses. Adenoviruses engineered to express hFc or Dll4–Fc were injected into the jugular vein of mice bearing subcutaneous C6 glioma, mouse mammary or HT1080 tumours. Adenoviruses provided systemic production of the engineered proteins.

Antibodies. Anti-Dll4 polyclonal antibodies were generated by immunizing rabbits against murine Dll4–hFc protein. Serum was depleted for antibodies with reactivity to human Fc and then used to stain tumour sections or treat tumour-bearing mice.

Reporter mice. Dll4 Lac/Z reporter mice¹⁷ generated using Velocigene technology³² were implanted with Lewis lung tumour cells. Tumours were stained with antibodies to CD31/Pecam-1 and/or β -galactosidase, or reacted with 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal), and counterstained with pyronin-Y.

In vitro assays. *In vitro* assays to determine the effect of Dll4–Fc and full length Dll4 on Notch signalling used confluent human umbilical vein endothelial cells (VEC Technologies) treated with Dll4–Fc protein (10 μ g ml^{–1}) for 2 to 8 h. RNA was extracted and analysed by Taqman, using probes and primers specific for human HES1, HEY2 and NRARP. In other experiments, human umbilical vein endothelial cells (50% confluent) were co-cultured with C6 glioma cells and assayed at 24 h.

Assaying hypoxia and vessel perfusion. To measure hypoxia and vessel perfusion, HypoxyProbe-1 (Chemicon; 60 mg kg^{–1}) was injected intraperitoneally one hour before sacrifice. Tumours were processed for histological analysis, and tumour sections were stained using anti-Hypoxyprobe antibody. To mark vessel perfusion, mice were injected through the jugular vein with biotinylated *Lycopersicon esculentum* lectin (100 μ g, Vector Laboratories). Lectin circulated for 3 min, and then tumours were subsequently stained for lectin bound to the endothelial cell surface³³.

Received 17 June; accepted 16 October 2006.

1. Folkman, J. The role of angiogenesis in tumor growth. *Semin. Cancer Biol.* **3**, 65–71 (1992).
2. Ferrara, N. Vascular endothelial growth factor as a target for anticancer therapy. *Oncologist* **9** (Suppl. 1), 2–10 (2004).
3. Rudge, J. S. *et al.* VEGF trap as a novel antiangiogenic treatment currently in clinical trials for cancer and eye diseases, and VelociGene-based discovery of the next generation of angiogenesis targets. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 411–418 (2005).
4. Holash, J. *et al.* VEGF-Trap: a VEGF blocker with potent antitumor effects. *Proc. Natl Acad. Sci. USA* **99**, 11393–11398 (2002).
5. Hurwitz, H. *et al.* Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med.* **350**, 2335–2342 (2004).
6. Laskin, J. J. & Sandler, A. B. First-line treatment for advanced non-small-cell lung cancer. *Oncology* **19**, 1671–6; discussion 1678–80 (2005).
7. Casanovas, O., Hicklin, D. J., Bergers, G. & Hanahan, D. Drug resistance by evasion of antiangiogenic targeting of VEGF signaling in late-stage pancreatic islet tumors. *Cancer Cell* **8**, 299–309 (2005).
8. Jain, R. K., Duda, D. G., Clark, J. W. & Loeffler, J. S. Lessons from phase III clinical trials on anti-VEGF therapy for cancer. *Nature Clin. Pract. Oncol.* **3**, 24–40 (2006).
9. Kerbel, R. S. *et al.* Possible mechanisms of acquired resistance to anti-angiogenic drugs: implications for the use of combination therapy approaches. *Cancer Metastasis Rev.* **20**, 79–86 (2001).
10. Yancopoulos, G. D. *et al.* Vascular-specific growth factors and blood vessel formation. *Nature* **407**, 242–248 (2000).
11. Jain, R. K. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* **307**, 58–62 (2005).

12. Carmeliet, P. Angiogenesis in life, disease and medicine. *Nature* **438**, 932–936 (2005).
13. Shawber, C. J. & Kitajewski, J. Notch function in the vasculature: insights from zebrafish, mouse and man. *Bioessays* **26**, 225–234 (2004).
14. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
15. Gridley, T. Notch signaling during vascular development. *Proc. Natl Acad. Sci. USA* **98**, 5377–5378 (2001).
16. Duarte, A. *et al.* Dosage-sensitive requirement for mouse Dll4 in artery development. *Genes Dev.* **18**, 2474–2478 (2004).
17. Gale, N. W. *et al.* Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development. *Proc. Natl Acad. Sci. USA* **101**, 15949–15954 (2004).
18. Krebs, L. T. *et al.* Haploinsufficient lethality and formation of arteriovenous malformations in Notch pathway mutants. *Genes Dev.* **18**, 2469–2473 (2004).
19. Mailhos, C. *et al.* Delta4, an endothelial specific notch ligand expressed at sites of physiological and tumor angiogenesis. *Differentiation* **69**, 135–144 (2001).
20. Patel, N. S. *et al.* Up-regulation of delta-like 4 ligand in human tumor vasculature and the role of basal expression in endothelial cell function. *Cancer Res.* **65**, 8690–8697 (2005).
21. Hicks, C. *et al.* A secreted Delta1-Fc fusion protein functions both as an activator and inhibitor of Notch1 signaling. *J. Neurosci. Res.* **68**, 655–667 (2002).
22. Taylor, K. L., Henderson, A. M. & Hughes, C. C. Notch activation during endothelial cell network formation *in vitro* targets the basic HLH transcription factor HESR-1 and downregulates VEGFR-2/KDR expression. *Microvasc. Res.* **64**, 372–383 (2002).
23. Iso, T., Kedes, L. & Hamamori, Y. HES and HERP families: multiple effectors of the Notch signaling pathway. *J. Cell. Physiol.* **194**, 237–255 (2003).
24. Shawber, C. J., Das, I., Francisco, E. & Kitajewski, J. Notch signaling in primary endothelial cells. *Ann. NY Acad. Sci.* **995**, 162–170 (2003).
25. Karsan, A. The role of notch in modeling and maintaining the vasculature. *Can. J. Physiol. Pharmacol.* **83**, 14–23 (2005).
26. Lamar, E. *et al.* Nrarp is a novel intracellular component of the Notch signaling pathway. *Genes Dev.* **15**, 1885–1899 (2001).
27. Krebs, L. T., Deftos, M. L., Bevan, M. J. & Gridley, T. The *Nrarp* gene encodes an ankyrin-repeat protein that is transcriptionally regulated by the Notch signaling pathway. *Dev. Biol.* **238**, 110–119 (2001).
28. Krneta, J. *et al.* Dissociation of angiogenesis and tumorigenesis in follistatin- and activin-expressing tumors. *Cancer Res.* **66**, 5686–5695 (2006).
29. Lee, C. G. *et al.* Anti-vascular endothelial growth factor treatment augments tumor radiation response under normoxic or hypoxic conditions. *Cancer Res.* **60**, 5565–5570 (2000).
30. Jain, R. K. Tumor angiogenesis and accessibility: role of vascular endothelial growth factor. *Semin. Oncol.* **29**, 3–9 (2002).
31. Jain, R. K. Antiangiogenic therapy for cancer: current and emerging concepts. *Oncology* **19**, 7–16 (2005).
32. Valenzuela, D. M. *et al.* High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nature Biotechnol.* **21**, 652–659 (2003).
33. Thurston, G., Baluk, P., Hirata, A. & McDonald, D. M. Permeability-related changes revealed at endothelial cell borders in inflamed venules by lectin binding. *Am. J. Physiol.* **271**, H2547–H2562 (1996).
34. Liu, Z. J. *et al.* Inhibition of endothelial cell proliferation by Notch1 signaling is mediated by repressing MAPK and PI3K/Akt pathways and requires MAML1. *FASEB J.* **20**, 1009–1011 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the following Regeneron colleagues: Y. Wei for gene expression analysis, A. Adler, A. Rafique, B. Li, H. Huang, E. Pasnikowski, J. McClain, E. Burova, D. Hylton, P. Burfeind and J. Griffiths for technical assistance, S. Staton for assistance with graphics, and S. Wiegand, I. Lobov, T. Daly, S. Davis, E. Ioffe, J. Holash and J. Rudge for scientific input.

Author Contributions I. N.-T. directed and helped perform tumour experiments, generation of tumour lines, immunohistochemical staining, and data analysis. C.D. directed, helped perform, and analysed *in vitro* experiments. N.J.P. helped develop protein reagents and biochemical assays. S.C. performed and helped analyse tumour experiments and construction of tumour cell lines. P.B. performed and helped analyse immunohistochemical studies. N.W.G. helped perform and analyse experiments with gene-targeted mice. H.C.L. helped perform and analyse gene expression studies. G.D.Y. helped analyse and interpret results. G.T. helped design experiments, analyse data and interpret results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to G.T. (Gavin.Thurston@Regeneron.com).

ARTICLES

A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis

Laura J. Niedernhofer^{1,4}, George A. Garinis¹, Anja Raams¹, Astrid S. Lalai¹, Andria Rasile Robinson⁴, Esther Appeldoorn¹, Hanny Odijk¹, Roos Oostendorp¹, Anwaar Ahmad⁴, Wibeke van Leeuwen², Arjan F. Theil¹, Wim Vermeulen¹, Gijsbertus T. J. van der Horst¹, Peter Meinecke⁵, Wim J. Kleijer³, Jan Vijg⁶, Nicolaas G. J. Jaspers¹ & Jan H. J. Hoeijmakers¹

XPF-ERCC1 endonuclease is required for repair of helix-distorting DNA lesions and cytotoxic DNA interstrand crosslinks. Mild mutations in *XPF* cause the cancer-prone syndrome xeroderma pigmentosum. A patient presented with a severe *XPF* mutation leading to profound crosslink sensitivity and dramatic progeroid symptoms. It is not known how unrepaired DNA damage accelerates ageing or its relevance to natural ageing. Here we show a highly significant correlation between the liver transcriptome of old mice and a mouse model of this progeroid syndrome. Expression data from *XPF-ERCC1*-deficient mice indicate increased cell death and anti-oxidant defences, a shift towards anabolism and reduced growth hormone/insulin-like growth factor 1 (IGF1) signalling, a known regulator of lifespan. Similar changes are seen in wild-type mice in response to chronic genotoxic stress, caloric restriction, or with ageing. We conclude that unrepaired cytotoxic DNA damage induces a highly conserved metabolic response mediated by the IGF1/insulin pathway, which re-allocates resources from growth to somatic preservation and life extension. This highlights a causal contribution of DNA damage to ageing and demonstrates that ageing and end-of-life fitness are determined both by stochastic damage, which is the cause of functional decline, and genetics, which determines the rates of damage accumulation and decline.

Numerous progeroid syndromes are caused by defects in the cellular response to DNA damage, including Cockayne syndrome, Werner syndrome, ataxia telangiectasia and trichothiodystrophy¹, suggesting that defective genome maintenance contributes to ageing. However, the value of progeroid syndromes for understanding normal human ageing is controversial^{2–4}, primarily because symptoms are tissue-specific. The segmental nature of these progeroid syndromes is consistent with the disposable soma theory of ageing (which posits that ageing results from accumulated damage⁵), because damage is stochastic and each tissue has different requirements for the various repair mechanisms⁶. Nevertheless, the most consistent determinant of lifespan is the mitogenic growth hormone/IGF1 pathway^{7,8}. Prolonged dampening of the axis genetically or by caloric restriction promotes longevity, whereas persistent upregulation shortens life. Experimental evidence reconciling the apparently disparate mechanisms of progeroid syndromes and natural ageing is currently lacking.

Nucleotide excision repair (NER) is a multi-step 'cut and patch' mechanism that removes distorting lesions affecting one strand of DNA such as those resulting from ultraviolet (UV) radiation damage. Two subpathways exist: transcription-coupled NER (TC-NER) and global genome NER (GG-NER). TC-NER removes lesions that block RNA polymerases, rescuing transcription and preventing cell death. Defective TC-NER causes Cockayne syndrome and trichothiodystrophy⁹. GG-NER operates genome-wide, primarily preventing mutations. Defective GG-NER causes the cancer-prone syndrome

xeroderma pigmentosum⁹. Xeroderma pigmentosum patients (complementation groups XP-A to XP-G) have over a 1,000-fold increased risk of skin cancer and a 10-fold increased risk of other tumours⁹. The contrasting phenotypes of Cockayne syndrome, trichothiodystrophy and xeroderma pigmentosum suggest that the cellular response to DNA damage dictates outcome: cell death/senescence accelerates ageing whereas cell survival (with mutations) promotes cancer¹⁰.

XPF-ERCC1 is an endonuclease required for NER¹¹. XPF is catalytic¹² and ERCC1 is essential for DNA binding¹³. Uniquely, this complex is also required for DNA interstrand crosslink (ICL) repair¹⁴. ICLs link both strands of DNA, preventing transcription and replication, and hence are extremely cytotoxic. Remarkably, XP-F patients have mild xeroderma pigmentosum¹¹, residual repair and only subtle *XPF* mutations. No patients with mutations in *ERCC1* are reported. Because complete inactivation of GG-NER or TC-NER is compatible with life, this suggests that the additional functions of XPF-ERCC1, including ICL repair, are essential. Here we describe a progeroid syndrome caused by a severe mutation in *XPF* and evidence from a mouse model of the disease that the growth hormone/IGF1 hormonal axis, a known regulator of lifespan, is suppressed in response to cytotoxic DNA damage.

A progeroid syndrome due to mutation of *XPF*

A boy aged 15 presented with frequent sunburns and a unique combination of progeroid symptoms (Supplementary Fig. 1 and

¹Center for Biomedical Genetics Medical Genetic Center Department of Cell Biology and Genetics, and ²Department of Experimental Radiology, and ³Department of Clinical Genetics, Erasmus Medical Center, PO Box 1738 3000 DR Rotterdam, The Netherlands. ⁴University of Pittsburgh Cancer Institute, Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, 5117 Centre Avenue, Pittsburgh, Pennsylvania 15213, USA. ⁵Abteilung für Medizinische Genetik, Altonaer KinderKrankenhaus, Bleickenallee 38, 22763 Hamburg, Germany. ⁶The Buck Institute for Age Research, 8001 Redwood Boulevard, Novato, California 94945, USA.

Supplementary Information 1). Inborn photosensitivity is indicative of defective NER. Indeed, poor survival after UV irradiation, reduced UV-radiation-induced DNA repair synthesis and impaired RNA synthesis recovery after UV irradiation of the patient's fibroblasts (the XPF-ERCC1 (XFE) fibroblasts; Fig. 1a, b; Supplementary Fig. 2 and Supplementary Information 2 Methods) indicated an almost complete absence of GG-NER and TC-NER, characteristic of both xeroderma pigmentosum and Cockayne syndrome. However, the neurologic, hepatobiliary, musculoskeletal and haematopoietic symptoms clearly discriminate this syndrome from xeroderma pigmentosum, Cockayne syndrome or combined xeroderma pigmentosum-Cockayne syndrome. Complementation analysis indicated that XFE cells are defective in XPF (Supplementary Fig. 2). This was unanticipated: all XP-F patients have substantial residual NER and mild xeroderma pigmentosum^{15,16}. XFE complementary DNA revealed a G→C transversion at position 458 in *XPF*, predicting a non-conservative substitution of a highly conserved arginine (R153P) (Supplementary Fig. 2; GenBank Accession number NM_005236). The patient's genomic DNA demonstrated homozygosity for this mutation. R153 resides in a domain harbouring helicase motifs¹³ and a leucine-rich region frequently involved in protein interactions. XPF and ERCC1 levels were lower in XFE cells than cells from XP-F patients (Fig. 1c, d), but detectable. XFE fibroblasts were exquisitely sensitive to ICL damage (Fig. 1e), confirming a role for XPF-ERCC1 in ICL resistance, distinct from that in NER. The unique clinical and cellular parameters define a novel disorder that we term XFE progeroid syndrome.

Ercc1^{-/-} mice model XFE progeroid syndrome

In mice, ERCC1- and XPF-null mutants are viable¹⁷⁻¹⁹ (Supplementary Table 1), but their severe phenotype is quite distinct from NER deficiency²⁰. Embryonic and early post-natal development of

Ercc1^{-/-} mice is mildly retarded, but growth arrests dramatically in the second week, typically culminating in death by four weeks (Fig. 2a, b and Supplementary Fig. 3a, b). *Ercc1*^{-/-} mice show ageing-like skin, liver and bone marrow abnormalities^{17,18,21}. We additionally identified dystonia and progressive ataxia (both indicative of neurodegeneration), renal insufficiency, sarcopenia, kyphosis and, at the cellular level, premature replicative senescence and sensitivity to oxidative stress (Fig. 2c, d and Supplementary Figs 3c and 4)—all changes associated with advanced age.

Early postnatal development was normal in patient 'XFE'. The progeroid symptoms initiated in early prepubescence, resulted in death before sexual maturation and included an old, wizened appearance, weight loss, epidermal atrophy, visual and hearing loss, ataxia, cerebral atrophy, hypertension, liver dysfunction, anaemia, osteopaenia, kyphosis, sarcopaenia and renal insufficiency. There is a striking correlation between the human syndrome and the *Ercc1*^{-/-} mouse phenotype (Supplementary Table 2). Importantly, XPF is undetectable in *Ercc1*^{-/-} mouse tissue (Supplementary Fig. 3d), indicating destabilization of the complex²². Furthermore, like XFE,

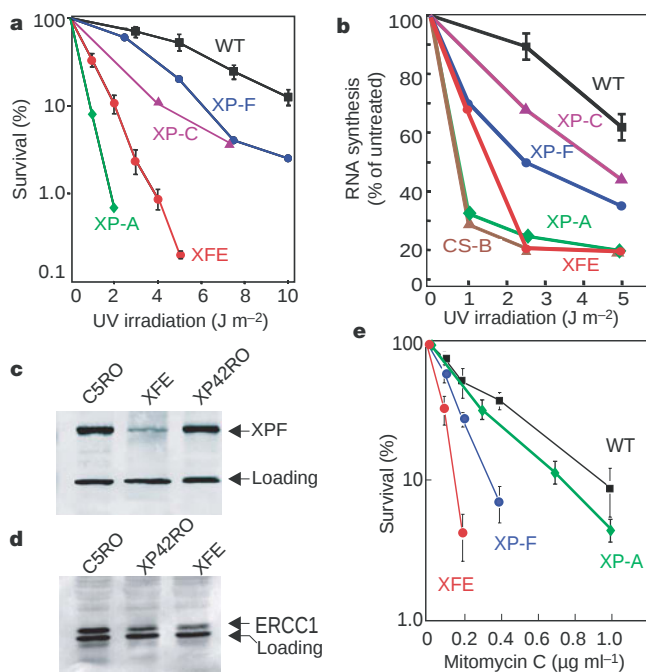


Figure 1 | Molecular characterization of progeroid patient 'XFE'.

a, Clonogenic survival assay measuring UV-radiation sensitivity of wild-type (WT), XFE and xeroderma pigmentosum patient primary fibroblasts (XP-F, XP-C and XP-A). **b**, RNA synthesis recovery after UV irradiation of patient fibroblasts. **c**, Immunodetection of XPF in nuclear extracts of patient fibroblasts (normal C5RO, mild xeroderma pigmentosum patient XP42RO and patient XFE). Cross-reacting bands demonstrate equal protein loading. **d**, Immunodetection of ERCC1 in the same samples. **e**, Clonogenic survival assay measuring sensitivity of patient fibroblasts to the crosslinking agent mitomycin C. Error bars (a, b, e) indicate s.e.m. of three experiments.

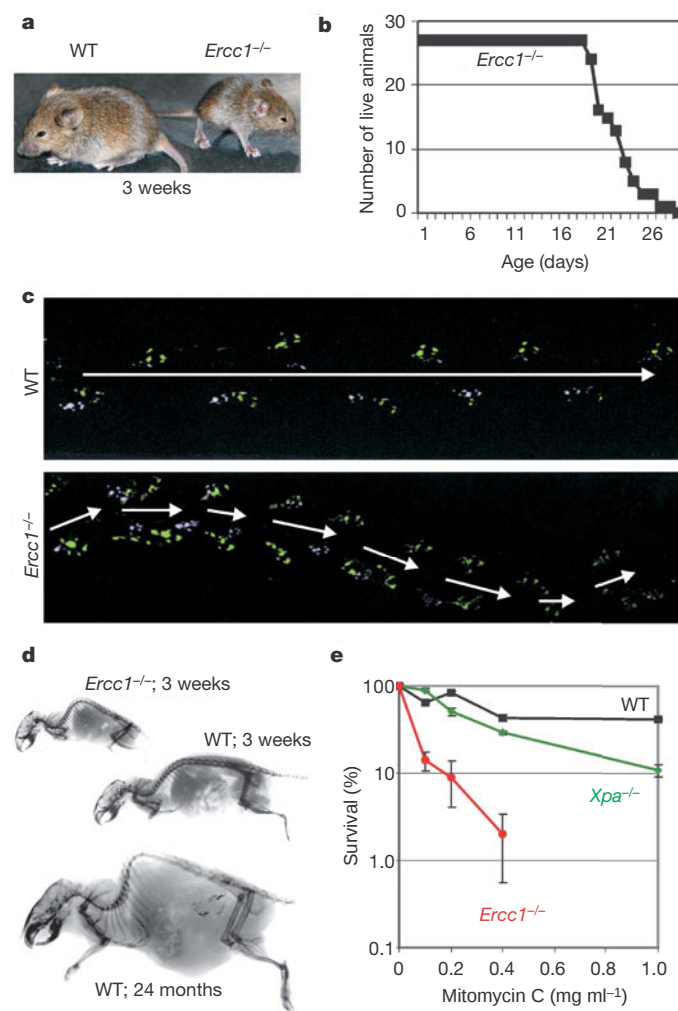


Figure 2 | Progeroid characteristics of *Ercc1*^{-/-} mice. **a**, *Ercc1*^{-/-} mouse and wild-type littermate at 3 weeks of age. **b**, Lifespan of *Ercc1*^{-/-} mice (*n* = 27). **c**, Footprint analysis of 3-week-old mice. Forepaws were painted purple, hind paws green. The arrows indicate gait trajectory. Fore- and hind-prints are not superimposed in mutant animals, a diagnostic criteria of ataxia. **d**, Radiographs demonstrating kyphosis in an aged wild-type and in an *Ercc1*^{-/-} mouse, but not in a young wild-type mouse. **e**, Clonogenic survival assay measuring the sensitivity of primary mouse embryonic fibroblasts to the crosslinking agent mitomycin C. Error bars indicate s.e.m. for three experiments, each averaging three replica platings of cells.

Ercc1^{-/-} cells are more sensitive to ICL damage than other NER-deficient cells (Fig. 2e). Collectively, these data establish *Ercc1*^{-/-} mice as an accurate model of the XFE progeroid syndrome.

Expression changes in *Ercc1*^{-/-} mouse liver

To investigate the cause of the progeroid features, we compared the entire transcriptome of *Ercc1*^{-/-} mouse liver with that of wild-type littermates at the age of 15 days, when the *Ercc1*^{-/-} mice reached maximal weight (Supplementary Fig. 3b), and had progeroid symptoms but only limited pathology. The liver was selected because it showed well-defined ageing-related changes and evidence for genotoxic stress (stabilized p53)¹⁷. Two-tailed, *t*-test analysis of variance of Affymetrix full genome arrays revealed 1,675 genes with significantly changed expression in *Ercc1*^{-/-} compared with wild-type liver ($\geq \pm 1.2$ -fold change and $P \leq 0.01$; Supplementary Table 3). Gene ontology classification (Supplementary Table 4) revealed: (1) global attenuation of the somatotroph, lactotroph and thyrotroph hormonal axes and other growth-promoting mechanisms; (2) downregulation of oxidative and carbohydrate metabolism; (3) upregulation of genes associated with glycogen synthesis but downregulation of glycogen phosphorylase, indicating a propensity to store not use glucose, an anticipated response to hyposomatotropism; (4) upregulation of fatty acid synthesis, the leptin receptor (*Lepr*) and peroxisome proliferator-activated receptor (*Ppar*)- γ and α , further suggesting an attempt to store energy; (5) upregulation of anti-oxidant and detoxification defences; (6) upregulation of DNA repair, suggesting increased damage load; and (7) upregulation of pro-apoptotic genes and downregulation of anti-apoptotic genes, indicating spontaneous hepatocellular toxicity. Remarkably, the majority of these metabolic changes are associated with extended lifespan in *Caenorhabditis elegans*²³.

Quantitative real-time PCR (qRT-PCR) of key genes in the pathways confirmed these findings in the liver, extended them to the kidney, and identified a trend towards increasing changes in expression with age (Fig. 3a, b), indicating a systemic, progressive process. The predicted downregulation of the growth hormone/IGF1 somatotroph axis in *Ercc1*^{-/-} mice explains their post-natal growth retardation. IGF1 is synthesized and secreted by the liver in response to growth hormone and is primarily responsible for its growth-promoting effects²⁴; thus it is a terminal read-out of the somatotroph axis. Circulating IGF1 levels in *Ercc1*^{-/-} mice were significantly lower than wild-type littermates (Fig. 3c; $P < 0.001$). These data predict a state of glucose storage rather than usage in *Ercc1*^{-/-} mice. In accordance, blood glucose levels were significantly lower than control littermates (Fig. 3d; $P < 0.001$). *Ercc1*^{-/-} mice were also hypoinsulinaemic (Fig. 3e; $P < 0.03$), consistent with insulin sensitivity and

elevated expression of *Ppar*- γ . These data validate the metabolic changes predicted by microarray analysis.

Suppression of the growth hormone axis in response to DNA damage

A primary defect in the hypothalamus or pituitary gland is not the cause of low IGF1 in *Ercc1*^{-/-} mice: immunohistochemistry of the anterior pituitary (which contains somatotrophs) revealed no pathology, normal numbers of growth-hormone-positive cells (Fig. 4a), and viable somatotrophs (Fig. 4b). Moreover, serum growth hormone of *Ercc1*^{-/-} mice was within the normal range (Fig. 4c) or tended to be elevated, as expected in response to low IGF1 and growth hormone receptor through feedback regulatory mechanisms. This raises the possibility that reduced IGF1 represents an adaptive response to DNA damage. Interestingly, chronic exposure of adult wild-type mice to subtoxic doses of the crosslinking agent mitomycin C suppressed serum IGF1 (Fig. 4d). In addition, qRT-PCR data indicated similar expression changes in the liver of mitomycin-C-exposed mice as seen in *Ercc1*^{-/-} mice (Supplementary Fig. 5). These results demonstrate that the systemic response in *Ercc1*^{-/-} mice, resulting from their DNA repair defect, is elicited in normal organisms that are challenged with genotoxic stress.

Parallels between XPF-ERCC1 deficiency and ageing

The somatotroph axis declines with age in mammals, including man^{8,25}. This prompted us to systematically compare broader expression changes in *Ercc1*^{-/-} and aged mice. The complete liver transcriptomes of 16-week-old and 130-week-old wild-type mice were compared with that of 8-week-old wild-type mice to identify expression changes in young adult (16 weeks) and aged (130 weeks) mice relative to juveniles (8 weeks) (Supplementary Table 5–7). Gene ontology classification of these gene sets indicated global suppression of the somatotroph axis, carbohydrate and oxidative metabolism, peroxisome biogenesis and ATP synthesis, but upregulation of immune and inflammatory responses and protein glycosylation in aged, but not young adult mice, concurring with previous studies in aged murine skeletal muscle and liver^{26,27}. Biological pathways that were similarly affected in *Ercc1*^{-/-} and aged wild-type mice included: the somatotroph axis, carbohydrate and oxidative metabolism, and peroxisome biogenesis. Apoptotic and anti-oxidant responses were largely unique to *Ercc1*^{-/-} mice, whereas inflammatory responses and protein glycosylation were unique to aged mice.

To quantify the similarity between *Ercc1*^{-/-} and aged mice, we asked how many of the 1,675 differentially expressed genes in the *Ercc1*^{-/-} liver transcriptome were altered in the same direction in aged mice? A 100% correlation yields a Spearman rank correlation

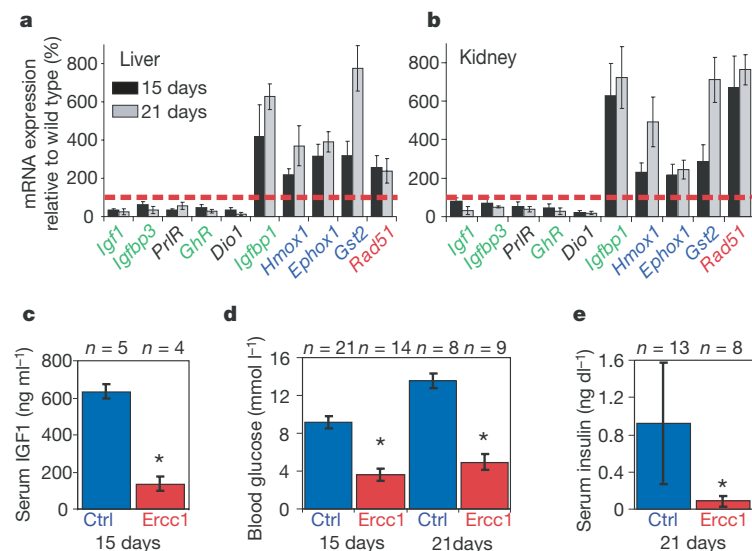


Figure 3 | Confirmation of microarray expression data by qRT-PCR and physiologic endpoints. **a**, qRT-PCR of liver messenger RNA levels of genes associated with the growth hormone/IGF1 axis (green), other hormonal axes (black), oxidant defence (blue), and DNA repair (red) in the liver of 15-day-old (black) and 21-day-old (grey) *Ercc1*^{-/-} mice relative to wild-type-littermates (red line). Each bar represents the average of three animals $\pm$ s.d. **b**, Similar analysis on the kidney. **c**, Serum IGF1 levels in *Ercc1*^{-/-} mice and littermate controls (Ctrl). Average values for 4–5 animals are plotted $\pm$ s.e.m. (* $P < 0.001$). **d**, Blood glucose levels of 15-day-old and 21-day-old *Ercc1*^{-/-} mice and control littermates. Average values for 8–21 animals are plotted $\pm$ s.e.m. (* $P < 0.001$). **e**, Serum insulin levels of 3-week-old *Ercc1*^{-/-} mice and control littermates. Average values for 8–13 animals are plotted $\pm$ s.e.m. (* $P < 0.03$).

coefficient ρ (r) of 1.0. The expression pattern of *Ercc1*^{-/-} mice had a significant degree of correlation with old mice ($r = 0.32$, $P \leq 0.0001$), but not young mice ($r = -0.03$). When the comparison was restricted to those biological themes that were significantly over-represented in the *Ercc1*^{-/-} mouse liver, there was an 86% correlation and within the somatotroph axis 95% (Supplementary Table 8). Therefore, despite the tremendous difference in age and genetic background, the transcriptional changes in *Ercc1*^{-/-} mice and old mice showed a highly significant correlation. Importantly, the liver transcriptomes of *Xpa*^{-/-} (nullizygous for the essential NER protein XPA) and *Csb*^{m/m} mice (homozygous for a mutation in *Csb* that leads to Cockayne syndrome), which are NER-deficient but not progeroid, did not significantly overlap with that of old mice, but the transcriptome of severely progeroid *Xpa*^{-/-};*Csb*^{m/m} mice did ($r = 0.44$; $P \leq 0.0001$)²⁸. Thus, we identified a common pattern of gene expression changes between aged and progeroid DNA-repair-deficient mice.

Cellular proliferation was drastically decreased in both *Ercc1*^{-/-} and aged mice (Fig. 5a), as predicted by microarray analysis. Further, senescent, polyploid hepatocytes, a hallmark of ageing²⁹, were prominent in both the *Ercc1*^{-/-} and aged mice, but not immature animals (Fig. 5a). Apoptotic cells were dramatically increased in *Ercc1*^{-/-} mouse liver and to a lesser extent in aged mice (Fig. 5b). Lysochrome staining revealed triglyceride accumulation in the liver of *Ercc1*^{-/-} and aged mice, which is consistent with decreased oxidative

metabolism (Fig. 5c). Steatosis was more pronounced in 21- than 15-day-old *Ercc1*^{-/-} mice, indicating a progressive process. Finally, IGFBP1—a protein overexpressed in response to liver injury, fasting or hypoinsulinemia³⁰—was extremely elevated in *Ercc1*^{-/-} and aged-mouse liver compared with that of young wild-type mice (Fig. 5d). These data confirm that the diverse metabolic changes predicted by expression analysis are similarly altered in *Ercc1*^{-/-} and aged mice.

A model connecting DNA damage, the growth hormone axis and ageing

Different mutations in *XPF* result in distinct clinical outcomes: either cancer as in xeroderma pigmentosum, or progeroid symptoms as in XFE syndrome. One explanation is that the R153 XFE mutation, compromising both NER and ICL repair, results primarily in cell death and senescence in response to DNA damage. This suppresses carcinogenesis^{31,32} but enhances ageing^{33–35}. In contrast, the milder NER defect in classic XP-F patients causes less cell death, allowing mutation accumulation and consequently cancer (elaborated in Supplementary Information 3). The specific sensitivity of XPF–ERCC1-deficient cells to crosslink damage, make ICLs a likely candidate for contributing to the unique phenotype of the XFE progeroid syndrome. Although spontaneous ICLs escape detection by current technology, several abundant endogenous compounds, many of which are by-products of lipid peroxidation, are known to crosslink DNA^{36,37}. Intriguingly, IGFBP1, which is extremely elevated in the *Ercc1*^{-/-} mouse liver (Fig. 5d), is strongly induced in rodents exposed to the crosslinking agent cisplatin³⁸ or fed a diet rich in polyunsaturated fatty acids³⁹, which promotes lipid peroxidation⁴⁰. Our data indicate that accumulation of nuclear DNA damage causes many of the pathophysiologic and metabolic changes associated with ageing, probably through increased cell death or senescence, without mutations and telomere loss^{41,42}. This is consistent with the damage accumulation theory of ageing^{6,43} and predicts that cytotoxic genotoxins used in adjuvant chemotherapy for cancer may promote ageing⁴⁴.

A key observation was that the somatotroph axis is suppressed when DNA damage is increased either genetically or chemically. This result is consistent with the observation that genetic deletion of SIRT6, a protein ADP ribosyl-transferase that positively affects lifespan and genome integrity, also suppresses IGF1⁴⁵. In addition, overexpression of the short isoform of p53 in mice impairs genome maintenance and IGF1 signalling³⁴. Together these studies create an inexorable link between genome maintenance and the somatotroph axis.

Direct genetic disruption of the growth hormone axis in mice leads to metabolic changes, including low serum glucose and insulin, improved insulin sensitivity, increased expression of *Ppar-γ*, decreased β -oxidation of fatty acids, increased gluconeogenesis, and enhanced anti-oxidant defenses and stress resistance⁸. As a consequence, these mice have reduced growth, body weight and fertility, yet increased lifespan and delayed onset of age-associated morbidities⁸. Caloric restriction causes rapid and reversible transcriptional reprogramming resulting in decreased insulin/IGF1 signalling with similar downstream effects⁴⁶. Our data indicate a near-identical pattern of metabolic and growth changes in *Ercc1*^{-/-} mice. This supports the existence of a common stress response, mediated by the insulin pathway, providing self-protection in the face of diverse stressors—a phenomenon termed hormesis, with reference to caloric restriction⁴⁷.

Multiple types of cellular and molecular damage accumulate with age⁶, and could trigger the same stress response. This likelihood is supported by the highly significant correlation between the transcriptomes of *Ercc1*^{-/-} mice and old mice, as well as the fact that growth hormone/IGF1 signalling decreases with age in mammals⁴⁸. On this basis, we propose the following model to explain how DNA damage (and probably other types of molecular damage) contributes to ageing (Fig. 5e). Damage accumulates with time as a consequence of exposure to endogenous and environmental genotoxins and incomplete repair. The DNA damage triggers a stress response either

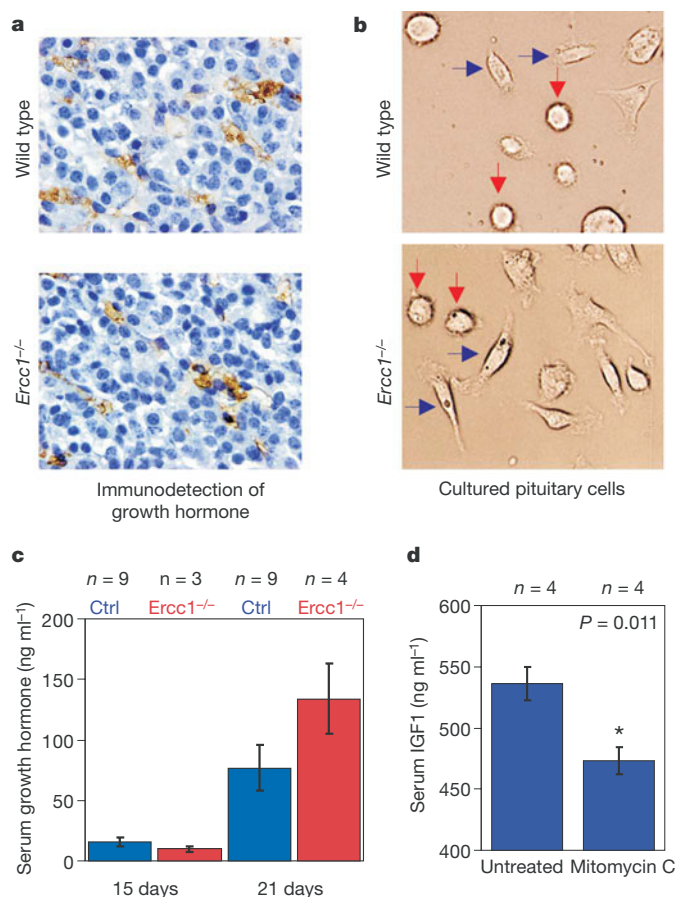


Figure 4 | Growth-hormone-IGF1 suppression is a normal physiological response to DNA damage. **a**, Immunodetection of growth hormone (brown) in anterior pituitary of wild-type and *Ercc1*^{-/-} mice. **b**, Cultured cells from the pituitary of wild-type and *Ercc1*^{-/-} mice indicating a normal distribution and viability of somatotrophs (blue arrows) and lactotrophs (red arrows). **c**, Serum growth hormone levels of 15- and 21-day-old *Ercc1*^{-/-} mice (red) and littermate controls (blue) (non-significant; $P = 0.4$, 15 days; $P = 0.12$, 21 days). **d**, Serum IGF1 levels of adult wild-type mice chronically exposed to mitomycin C (0.1 mg kg⁻¹ bimonthly) compared with untreated mice ($n = 4$ per group; * $P = 0.011$). Error bars, s.e.m.

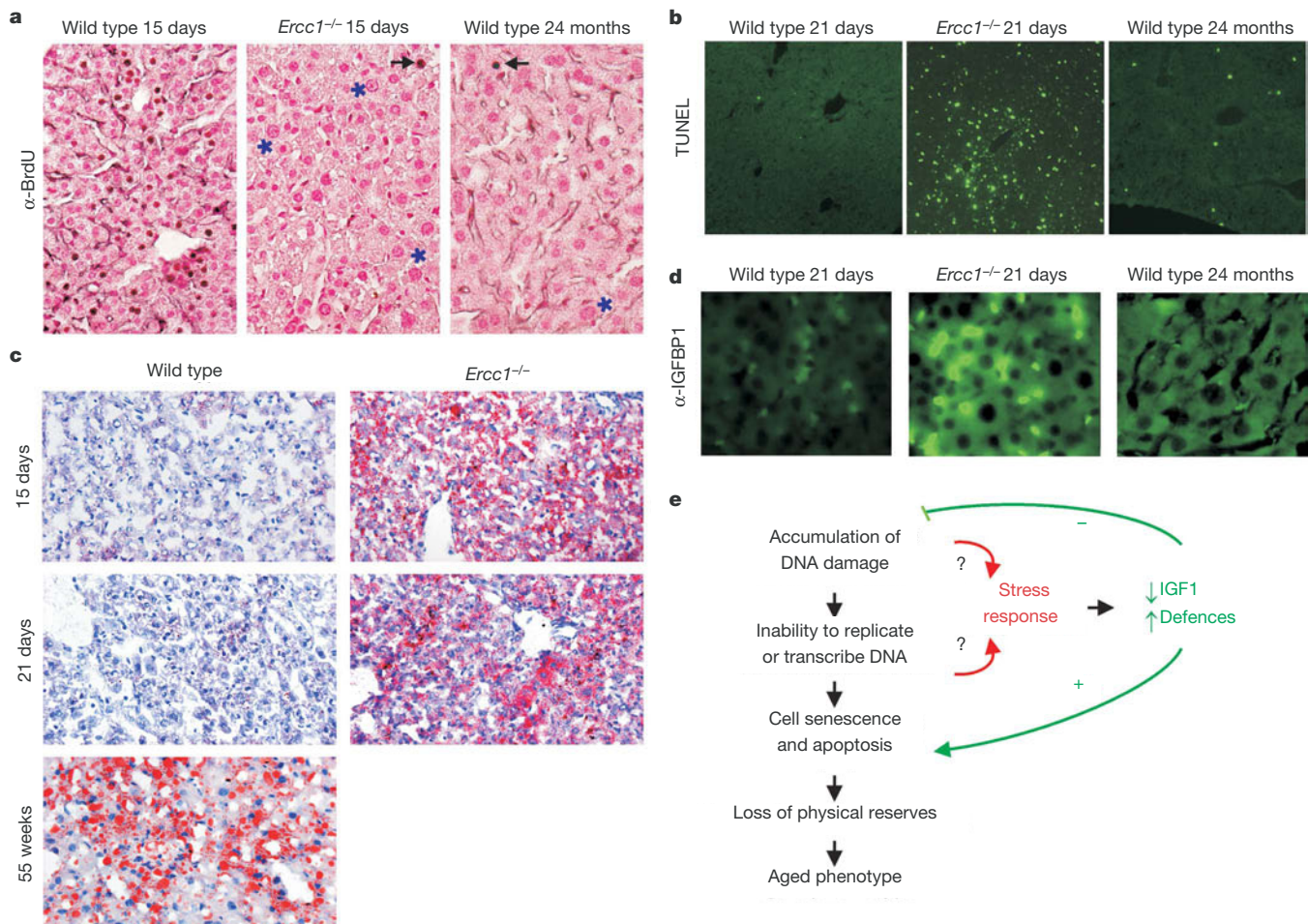


Figure 5 | Comparison of the physiological changes due to DNA repair defects and ageing. **a**, Detection of proliferating cells in liver of *Ercc1*^{-/-}, young and old wild-type mice by immunodetection of incorporated BrdU (brown nuclei and arrows). Large, polyploidy nuclei are apparent in the

directly or by interfering with transcription or replication. The response is systemic dampening of the growth hormone/IGF1 hormonal axis, through a mechanism that is, as yet, unknown, but is highly conserved⁶⁻⁸. This in turn leads to metabolic changes that shift energy usage from growth and proliferation to protective maintenance⁷, minimizing further damage, but contributing to apoptosis⁴⁹. Despite this protective response, organisms which continue to accumulate damage will inevitably age. Repair-deficient *Ercc1*^{-/-} mice continue to rapidly accumulate DNA damage and experience an early onset of degenerative processes. Aged organisms in which either damage prevention or repair mechanisms may already be compromised will continue to accumulate damage slowly, eventually succumbing to age-related morbidities and mortality. In both examples, the 'ageing process' is retarded by the insulin-mediated stress response.

This model reconciles two apparently disparate hypotheses of ageing: that ageing is genetically regulated⁷ and that ageing is a consequence of the accumulation of stochastic damage⁶. In fact, both are correct. Damage drives the functional decline that is associated with ageing; however, a highly conserved longevity assurance mechanism, mediated by the IGF1/insulin pathway, influences how rapidly damage accumulates and function is lost.

METHODS

Detailed methods are provided in Supplementary Information 2. Cell lines were established from skin biopsies of patients or from mouse embryos. Nucleotide excision repair was measured by a clonogenic survival assay after exposure of

Ercc1^{-/-} and aged mouse liver (asterisks). **b**, Measurement of apoptotic nuclei in liver by TUNEL assay. **c**, Detection of triglycerides in liver with the lysochrome Oil Red O. **d**, Immunodetection of IGFBP1 in the liver. **e**, Model of how DNA damage contributes to ageing (see text for details).

cells to UV irradiation (UV-C, 254 nm). Unscheduled DNA synthesis (a measure of GG-NER) was determined by quantifying incorporation of ³H-thymidine into genomic DNA after exposure of cells to 10 J m⁻² of UV-C radiation. RNA synthesis recovery (TC-NER) was measured as ³H-uridine incorporation after UV irradiation of cells. *XPF* was identified as the gene affected in patient XFE by fusing his cells with those of a patient from each complementation group of xeroderma pigmentosum and measuring unscheduled DNA synthesis. Both *XPF* cDNA and exons from gDNA isolated from XFE cells were sequenced. Crosslink sensitivity was determined by clonogenic survival assay following exposure to mitomycin C. Antibodies and ELISA kits used for immunodetection of *XPF*, *ERCC1*, IGF1, insulin, glucose, growth hormone, BrdU incorporation and IGFBP1; tissue staining; and TUNEL (TdT-mediated dUTP nick end labelling) reagents are elaborated in Supplementary Information 2.

Ercc1^{-/-} mice were generated in an F₁ hybrid background by crossing *Ercc1*^{+/-} mice. Genomic DNA was isolated from an ear punch and genotyped by PCR. Ataxia in mice was measured by analysing gait after painting fore- and hind-paws with different colours. X-rays were obtained with a CGR Senograph 500T instrument. Genome-wide expression profiles were determined for liver of 15-day-old *Ercc1*^{-/-} mice and wild-type littermates, as well as 8-, 16- and 130-week-old C57BL/6J mice using Affymetrix 430 V2.0 arrays. Expression changes were confirmed by qRT-PCR for 16 genes representing each of the biological themes significantly affected in *Ercc1*^{-/-} and aged mice. For chronic genotoxin exposure studies, adult wild-type mice (10 weeks old; *n* = 4) were administered 0.1 mg kg⁻¹ mitomycin C bimonthly for 10 weeks.

Received 11 September; accepted 20 November 2006.

- Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H. & Vijg, J. Aging and genome maintenance: lessons from the mouse? *Science* **299**, 1355–1359 (2003).
- Kipling, D., Davis, T., Ostler, E. L. & Faragher, R. G. What can progeroid syndromes tell us about human aging? *Science* **305**, 1426–1431 (2004).

3. Hasty, P. & Vijg, J. Accelerating aging by mouse reverse genetics: a rational approach to understanding longevity. *Aging Cell* **3**, 55–65 (2004).
4. Miller, R. A. Evaluating evidence for aging. *Science* **310**, 441–443; author reply 441–443 (2005).
5. Kirkwood, T. B. & Holliday, R. The evolution of ageing and longevity. *Proc. R. Soc. Lond. B* **205**, 531–546 (1979).
6. Kirkwood, T. B. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).
7. Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449–460 (2005).
8. Bartke, A. Minireview: role of the growth hormone/insulin-like growth factor system in mammalian aging. *Endocrinology* **146**, 3718–3723 (2005).
9. Bootsma, D., Kraemer, K. H., Cleaver, J. E. & Hoeijmakers, J. H. J. in *The Metabolic and Molecular Basis of Inherited Disease* (eds Scriver, C. R. et al.) 677–703 (McGraw-Hill, New York, 2001).
10. Mitchell, J. R., Hoeijmakers, J. H. & Niedernhofer, L. J. Divide and conquer: nucleotide excision repair battles cancer and ageing. *Curr. Opin. Cell Biol.* **15**, 232–240 (2003).
11. Sijbers, A. M. et al. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **86**, 811–822 (1996).
12. Enzlin, J. H. & Scharer, O. D. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J.* **21**, 2045–2053 (2002).
13. Tsodikov, O. V., Enzlin, J. H., Scharer, O. D. & Ellenberger, T. Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc. Natl Acad. Sci. USA* **102**, 11236–11241 (2005).
14. Niedernhofer, L. J. et al. The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Mol. Cell. Biol.* **24**, 5776–5787 (2004).
15. Matsumura, Y., Nishigori, C., Yagi, T., Imamura, S. & Takebe, H. Characterization of molecular defects in xeroderma pigmentosum group F in relation to its clinically mild symptoms. *Hum. Mol. Genet.* **7**, 969–974 (1998).
16. Sijbers, A. M. et al. Homozygous R788W point mutation in the XPF gene of a patient with xeroderma pigmentosum and late-onset neurologic disease. *J. Invest. Dermatol.* **110**, 832–836 (1998).
17. McWhir, J., Selfridge, J., Harrison, D. J., Squires, S. & Melton, D. W. Mice with DNA repair gene (ERCC-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. *Nature Genet.* **5**, 217–224 (1993).
18. Weeda, G. et al. Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Curr. Biol.* **7**, 427–439 (1997).
19. Tian, M., Shinkura, R., Shinkura, N. & Alt, F. W. Growth retardation, early death, and DNA repair defects in mice deficient for the nucleotide excision repair enzyme XPF. *Mol. Cell. Biol.* **24**, 1200–1205 (2004).
20. de Vries, A. et al. Increased susceptibility to ultraviolet-B and carcinogens of mice lacking the DNA excision repair gene XPA. *Nature* **377**, 169–173 (1995).
21. Prasher, J. M. et al. Reduced hematopoietic reserves in DNA interstrand crosslink repair-deficient Ercc1^{-/-} mice. *EMBO J.* **24**, 861–871 (2005).
22. Biggerstaff, M., Szymkowski, D. E. & Wood, R. D. Co-correction of the ERCC1, ERCC4 and xeroderma pigmentosum group F DNA repair defects *in vitro*. *EMBO J.* **12**, 3685–3692 (1993).
23. Hansen, M., Hsu, A. L., Dillin, A. & Kenyon, C. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. *PLoS Genet.* **1**, 119–128 (2005).
24. Carter, C. S., Ramsey, M. M. & Sonntag, W. E. A critical analysis of the role of growth hormone and IGF-1 in aging and lifespan. *Trends Genet.* **18**, 295–301 (2002).
25. Lombardi, G., Di Somma, C., Rota, F. & Colao, A. Associated hormonal decline in aging: is there a role for GH therapy in aging men? *J. Endocrinol. Invest.* **28**, 99–108 (2005).
26. Lee, C. K., Klopp, R. G., Weindrich, R. & Prolla, T. A. Gene expression profile of aging and its retardation by caloric restriction. *Science* **285**, 1390–1393 (1999).
27. Cao, S. X., Dhahbi, J. M., Mote, P. L. & Spindler, S. R. Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice. *Proc. Natl Acad. Sci. USA* **98**, 10630–10635 (2001).
28. van der Pluijm, I. et al. Impaired genome maintenance suppresses the GH/IGF1 axis in Cockayne syndrome mice. *PLoS Biol.* doi:10.1371/journal.pbio.0050002 (in the press).
29. Gupta, S. Hepatic polyploidy and liver growth control. *Semin. Cancer Biol.* **10**, 161–171 (2000).
30. Lee, P. D., Conover, C. A. & Powell, D. R. Regulation and function of insulin-like growth factor-binding protein-1. *Proc. Soc. Exp. Biol. Med.* **204**, 4–29 (1993).
31. Campisi, J. Aging, tumor suppression and cancer: high wire-act! *Mech. Ageing Dev.* **126**, 51–58 (2005).
32. Sonntag, W. E. et al. Adult-onset growth hormone and insulin-like growth factor I deficiency reduces neoplastic disease, modifies age-related pathology, and increases life span. *Endocrinology* **146**, 2920–2932 (2005).
33. Wyllie, F. S. et al. Telomerase prevents the accelerated cell ageing of Werner syndrome fibroblasts. *Nature Genet.* **24**, 16–17 (2000).
34. Maier, B. et al. Modulation of mammalian life span by the short isoform of p53. *Genes Dev.* **18**, 306–319 (2004).
35. Herbig, U., Ferreira, M., Condel, L., Carey, D. & Sedivy, J. M. Cellular senescence in aging primates. *Science* **311**, 1257 (2006).
36. Kasai, H., Iwamoto-Tanaka, N. & Fukada, S. DNA modifications by the mutagen glyoxal: adduction to G and C, deamination of C and GC and GA cross-linking. *Carcinogenesis* **19**, 1459–1465 (1998).
37. Niedernhofer, L. J., Daniels, J. S., Rouzer, C. A., Greene, R. E. & Marnett, L. J. Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. *J. Biol. Chem.* **278**, 31426–31433 (2003).
38. Huang, Q. et al. Assessment of cisplatin-induced nephrotoxicity by microarray technology. *Toxicol. Sci.* **63**, 196–207 (2001).
39. Ghoshal, A. K., Xu, Z., Wood, G. A. & Archer, M. C. Induction of hepatic insulin-like growth factor binding protein-1 (IGFBP-1) in rats by dietary n-6 polyunsaturated fatty acids. *Proc. Soc. Exp. Biol. Med.* **225**, 128–135 (2000).
40. Takahashi, Y., Kushi, M., Shinohara, K. & Ide, T. Activity and mRNA levels of enzymes involved in hepatic fatty acid synthesis and oxidation in mice fed conjugated linoleic acid. *Biochim. Biophys. Acta* **1631**, 265–273 (2003).
41. Zhu, X. D. et al. ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Mol. Cell* **12**, 1489–1498 (2003).
42. Dollé, M. E. et al. Increased genomic instability is not a prerequisite for shortened lifespan in DNA repair deficient mice. *Mutat. Res.* **596**, 22–35 (2006).
43. Lombard, D. B. et al. DNA repair, genome stability, and aging. *Cell* **120**, 497–512 (2005).
44. McCormick, R. E. Possible acceleration of aging by adjuvant chemotherapy: a cause of early onset frailty? *Med. Hypotheses* **67**, 212–215 (2006).
45. Mostoslavsky, R. et al. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315–329 (2006).
46. Spindler, S. R. Rapid and reversible induction of the longevity, anticancer and genomic effects of caloric restriction. *Mech. Ageing Dev.* **126**, 960–966 (2005).
47. Kirkwood, T. B. & Shanley, D. P. Caloric restriction, hormesis and life history plasticity. *Hum. Exp. Toxicol.* **19**, 338–339 (2000).
48. Muller, E. E., Locatelli, V. & Cocchi, D. Neuroendocrine control of growth hormone secretion. *Physiol. Rev.* **79**, 511–607 (1999).
49. Pinkston, J. M., Garigan, D., Hansen, M. & Kenyon, C. Mutations that increase the life span of *C. elegans* inhibit tumor growth. *Science* **313**, 971–975 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the National Institute of Aging Program, the Dutch Cancer Society, the Dutch Science Foundation (NWO) through the foundation of the Research Institute Diseases of the Elderly, as well as grants from SenterNovem IOP-Genomics, the NIH, the NIA Program Project, the NIEHS center, the EC, and Human Frontier Science Program. J.H.J.H. is chief scientific officer of DNAge. L.J.N. was supported by a postdoctoral fellowship from the American Cancer Society and subsequently by the NCI and The Ellison Medical Foundation, along with A.R.R. and A.A. We thank P. Nair, F. J. Calderon, R. B. Calder and D. Muñoz-Medellin of the Sam and Ann Barshop Center, University of Texas Health Science Center, for their contributions to the preliminary microarray analysis.

Author Contributions Human cells characterized by L.J.N., A.R., E.A., H.O., A.A., A.F.T., W.V. and N.G.J.J. Experimental analysis of mice by L.J.N., A.S.L., A.R.R. and R.O. Microarray analysis by L.J.N., G.A.G., G.T.J. v. d. H. and J.V. Radiography by W. v. L. Clinical services by P.M. and W.J.K. Manuscript prepared by L.J.N., G.A.G. and J.H.J.H.

Author Information The expression data for *Ercc1*^{-/-} mice are deposited in ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>), a public repository for microarray data, which stores Minimum Information for Microarray Experiments (MIAME)-compliant data in accordance with Microarray Gene Expression Data (MGED) recommendations. The accession number is E-MEXP-834. The accession number for the expression data on the aged mice is E-MEXP-839 (ref. 28). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H.J.H. (j.hoeijmakers@erasmusmc.nl).

LETTERS

A new γ -ray burst classification scheme from GRB 060614

N. Gehrels¹, J. P. Norris¹, S. D. Barthelmy¹, J. Granot², Y. Kaneko³, C. Kouveliotou⁴, C. B. Markwardt^{1,5}, P. Mészáros^{6,7}, E. Nakar⁸, J. A. Nousek⁶, P. T. O'Brien⁹, M. Page¹⁰, D. M. Palmer¹¹, A. M. Parsons¹, P. W. A. Roming⁶, T. Sakamoto^{1,12}, C. L. Sarazin¹³, P. Schady^{6,10}, M. Stamatikos^{1,12} & S. E. Woosley¹⁴

Gamma-ray bursts (GRBs) are known to come in two duration classes¹, separated at ~ 2 s. Long-duration bursts originate from star-forming regions in galaxies², have accompanying supernovae when these are near enough to observe and are probably caused by massive-star collapsars³. Recent observations^{4–10} show that short-duration bursts originate in regions within their host galaxies that have lower star-formation rates, consistent with binary neutron star or neutron star–black hole mergers^{11,12}. Moreover, although their hosts are predominantly nearby galaxies, no supernovae have been so far associated with short-duration GRBs. Here we report that the bright, nearby GRB 060614 does not fit into either class. Its ~ 102 -s duration groups it with long-duration GRBs, while its temporal lag and peak luminosity fall entirely within the short-duration GRB subclass. Moreover, very deep optical observations exclude an accompanying supernova^{13–15}, similar to short-duration GRBs. This combination of a long-duration event without an accompanying supernova poses a challenge to both the collapsar and the merging-neutron-star interpretations and opens the door to a new GRB classification scheme that straddles both long- and short-duration bursts.

The Burst Alert Telescope (BAT) onboard the Swift satellite¹⁶ detected GRB 060614 on 14 June 2006 at 12:43:48 UT. Using the initial positions from the onboard X-ray and ultraviolet/optical telescopes, the burst was subsequently located by ground- and space-based telescopes at the outskirts of a relatively nearby faint dwarf galaxy at a redshift of $z = 0.125$ (ref. 17). We do not find the suggestion^{18,19} of a chance alignment between a background GRB and foreground galaxy at $z = 0.125$ to be credible; the probability of the observed 0.5" offset between the GRB and the $z = 0.125$ galaxy occurring by chance is only 2×10^{-5} . Also, fits to the combined ultraviolet/optical and X-ray telescope spectra give $z < 1.3$ at the 99.99% confidence level, excluding the suggested¹⁸ location at $z > 1.4$ (see ref. 14 for additional evidence against a chance alignment).

The event's duration, $T_{90} = 102$ s (15–350 keV; where T_{90} is the time during which 90% of the event photons were collected), clearly places GRB 060614 in the long-duration burst category¹. The event was also at close proximity, so it became a prime candidate for a supernova search in its light curve, as indeed had been found in four individual cases in the past²⁰. The three accompanying papers^{13–15} report very tight upper limits—100 times lower than previous detections—in these searches. Thus far, strict supernova limits had been found only for the GRBs of the short-duration variety. So why

was there no supernova associated with the long-duration GRB 060614?

A close examination of the BAT light curve of GRB 060614 (Fig. 1) reveals a first short, hard-spectrum episode of emission (lasting 5 s) followed by an extended and somewhat softer episode (lasting ~ 100 s). The total energy content of the second episode is five times that of the first, with fluences of $(1.69 \pm 0.02) \times 10^{-5}$ erg cm⁻² and $(3.3 \pm 0.1) \times 10^{-6}$ erg cm⁻², respectively, in the 15–350 keV band. The recent discovery of a similar two-component emission structure in several Swift and the High Energy Transient Explorer 2 (HETE-2) spacecraft short-duration GRBs prompted Norris and Bonnell²¹ to search the Burst And Transient Source Experiment (BATSE) GRB database for similar events. They found three bursts with a bright softer emission component, spectrally similar to that of GRB 060614 (corresponding to roughly $\sim 1\%$ of the number of short-duration BATSE GRBs). Weaker soft components are found in four of 16 Swift and HETE-2 short bursts and 11 out of 130 short-duration GRBs observed by the Konus spacecraft²², which represents ~ 10 –25% of their short-duration GRB databases. It may be that the majority of short bursts have such soft components, since at present only two Swift short-duration GRBs (GRB 051221A and GRB 060313) are bright enough to allow low fractional limits to be set. We note that, regardless of their duration, almost all of these events with softer components have a common appearance: a 5–10-s gap between the first and second episodes and a hump-shaped second component.

Another method of distinguishing categories of GRBs is to compute the temporal lags between their light curve features in different energy bands^{21,23,24}. We introduce a new variant on this approach by including short bursts in a plot of peak luminosity (L_{peak}) versus lag (t_{lag}); the redshift measurements from the past year for a number of short-duration Swift GRBs now makes the comparison possible. There is an anti-correlation between t_{lag} and L_{peak} for long bursts as shown in Fig. 2. In contrast, short bursts have small t_{lag} and small L_{peak} and occupy a separate area of parameter space from long bursts. The lag for GRB 060614 for the first 5 s is 3 ± 6 ms. Thus, in spite of its long duration, GRB 060614 falls into the same region of the lag–luminosity plot as do short bursts. Moreover, we were able to accurately calculate a lag for the softer second episode, owing to its spiky light curve. We find $t_{\text{lag}} = 3 \pm 9$ ms, consistent with the short, hard episode, and thus indicating a similar origin. The physical cause of lags in GRBs is not yet well understood, but it has been suggested²³ that bursts with smaller lags have more relativistic outflows.

¹NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ²KIPAC, Stanford University, PO Box 20450, MS 29, Stanford, California 94309, USA. ³USRA, NSSTC, ⁴NASA/Marshall Space Flight Center, NSSTC, VP-62, 320 Sparkman Drive Huntsville, Alabama 35805, USA. ⁵Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. ⁶Department of Astronomy and Astrophysics, ⁷Department of Physics, Pennsylvania State University, Pennsylvania 16802, USA. ⁸Theoretical Astrophysics, California Institute of Technology, MS 130-33, Pasadena, California 91125, USA. ⁹Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. ¹⁰Mullard Space Science Laboratory, University College London, Dorking, RH5 6NT, UK. ¹¹Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ¹²Oak Ridge Associated Universities, PO Box 117, Oak Ridge, Tennessee 37831-0117, USA. ¹³Department of Astronomy, University of Virginia, Charlottesville, Virginia 22904-4325, USA. ¹⁴Department of Astronomy and Astrophysics, University of California at Santa Cruz, Santa Cruz, California 95064, USA.

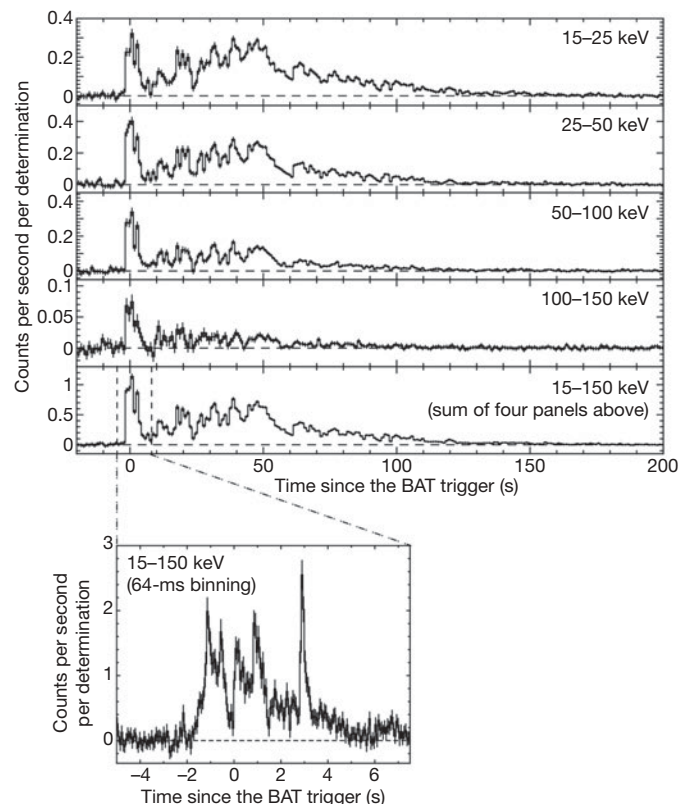


Figure 1 | The light curve of GRB 060614 as observed with the BAT. The mask-tagged BAT light curve is shown in four energy bands in the top panels, and these are summed in the bottom panel. The inset gives an expanded view of the first episode. There is a hint of a 9-s periodicity between 7 and 50 s, but it is not statistically significant. The spectrum is well fitted by a power law in both intervals with a photon index of 1.63 ± 0.07 in the first 5 s and a softer index of 2.02 ± 0.04 in the following 120 s. The burst was also detected with Konus-WIND²⁹ with $E_{\text{peak}} = 302 (-83, +214)$ keV in the first 5 s. The total energy radiated, assuming isotropic emission, E_{iso} , in the 1 keV–10 MeV range in the GRB restframe for the first 5 s is 1.8×10^{50} erg. We note that the $E_{\text{peak}} - E_{\text{iso}}$ value for the short episode falls far away from the Amati relationship for long-duration bursts. The spacecraft repointed the X-ray and ultraviolet/optical telescopes at the burst in 90 s. The X-ray emission (0.3–10 keV) was initially among the brightest until now at 5×10^{-8} erg cm⁻² s⁻¹ at 100 s. The afterglow light curves suggest a jet break at ~ 1.5 days, implying a narrow jet occupying $\leq 10^{-2}$ of the total solid angle and suggesting an energy of $\sim 10^{49}$ erg in the relativistic jet, roughly comparable to that inferred for Swift short-duration GRBs.

It is thus difficult to determine unambiguously which category GRB060614 falls into. It is a long-duration event by the traditional definition, but it lacks an associated supernova as observed in other nearby long-duration GRBs, and it shows morphological similarities with Swift short bursts. Further, the complex structure of the second component and lag similar to the first component indicates that the extended episode is due to the long-lived activity of the burst's central engine rather than due to the onset of the afterglow emission. However, its spectral lag and low luminosity make it distinct from other long-duration GRBs and place it instead with short bursts. Perhaps the odd characteristics of GRB 060614 point to a new subclass with a different physical origin identified by a third GRB property, namely spectral lag. However, we first must ask what would be needed to fit such a class of events into either the collapsar or merger models.

Short, hard bursts with continuing activity resembling long, soft bursts are a prediction of some massive star models²⁵, favoured when the pre-explosive mass loss is unusually high and the jet energy low. The lack of a supernova requires that most production of ^{56}Ni is avoided during the star collapse. This might be possible if the disk

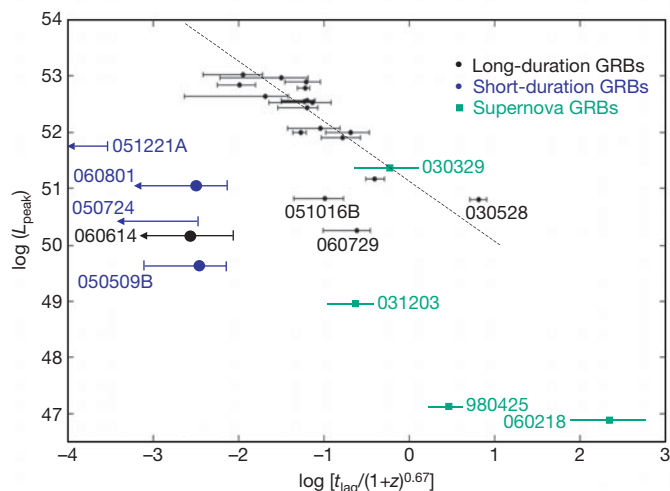


Figure 2 | Spectral lag as a function of peak luminosity showing GRB 060614 in the region of short-duration GRBs. The lags and peak luminosities are corrected to the source frame of the GRB. The lags are defined as the time difference between light curve structures in the 50–100 keV and 15–25 keV channels. The energy correction (K-correction) to the source frame for the lag is approximately $(1+z)^{0.33}$, making the total energy correction plus time dilation correction $(1+z)^{-0.67}$. The data points labelled as long bursts are from Swift/BAT, with the exception of GRB 030528 which is a very long-lagged HETE-2 burst. The dashed line illustrates the lag–luminosity correlation for long bursts. Outliers are bursts with unusual properties: GRB 060729 has an extremely long-lived afterglow and GRB 051016B has extremely soft prompt emission. The blue data points for short bursts are from Swift/BAT. The peak luminosity was calculated for the short bursts and GRB 060614 using 64-ms binning. In green are the four nearby long-duration GRBs with associated supernovae. Three of the four supernovae-associated underluminous nearby GRBs (GRB 980425, GRB 031203 and GRB 060218) fall below the long-burst correlation, while the only supernova-associated GRB with normal luminosity (GRB 030329) falls near the line, further suggesting that it is a normal GRB (the only normal burst with a firm supernova association). None of the supernova-associated GRBs falls near the short grouping. Note that our detailed calculation of the lag of GRB 030329 gives a larger value here than previously found³⁰ in a higher energy band. All error bars are 1σ .

wind is weak and if ^{56}Ni that had been produced initially fell back onto the collapsed core in a weak supernova explosion. Nevertheless, producing less than 10^{-3} solar masses of ^{56}Ni is a challenge for the collapsar model²⁶. Another possibility is a failed supernova; the small observed lag may imply a higher-velocity outflow with insufficient energy deposited in the star for it to explode. In the neutron-star merger model, the duration of GRB 060614 is a problem. The duration is set by the lifetime of the accretion disk around the newly born black hole and is expected²⁷ to be ~ 0.1 s. The accretion during a black hole–neutron star merger might be longer if, as suggested by simulations²⁸, a significant amount of mass is ejected to a bound orbit and falls back into the black hole on timescales larger than 1 s. We conclude that although GRB 060614 may be just another odd event in the motley-crew collection of GRBs, it is more probably one of the first-established members of an emerging class of events that includes both short- and long-duration GRBs.

Received 30 August; accepted 20 October 2006.

1. Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
2. Fruchter, A. *et al.* Long γ -ray bursts and core-collapse supernovae have different environments. *Nature* **441**, 463–468 (2006).
3. Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
4. Gehrels, N. *et al.* A short gamma-ray burst apparently associated with an elliptical galaxy at $z = 0.225$. *Nature* **437**, 851–854 (2005).
5. Fox, D. B. *et al.* The afterglow of GRB 050709 and the nature of the short-hard gamma-ray bursts. *Nature* **437**, 845–851 (2005).

6. Villasenor, J. S. *et al.* Discovery of the short gamma-ray burst GRB 050709. *Nature* **437**, 855–878 (2005).
7. Hjorth, J. *et al.* The optical afterglow of the short gamma-ray burst GRB 050709. *Nature* **437**, 859–861 (2005).
8. Barthelmy, S. D. *et al.* An origin for short gamma-ray bursts unassociated with current star formation. *Nature* **438**, 994–996 (2005).
9. Berger, E. *et al.* The afterglow and elliptical host galaxy of the short gamma-ray burst GRB 050724. *Nature* **438**, 988–990 (2005).
10. Bloom, J. S. *et al.* Closing in on a short-hard burst progenitor: constraints from early-time optical imaging and spectroscopy of a possible host galaxy of GRB 050509b. *Astrophys. J.* **638**, 354–368 (2006).
11. Paczynski, B. Gamma-ray bursters at cosmological distances. *Astrophys. J.* **308**, L43–L46 (1986).
12. Eichler, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars. *Nature* **340**, 126–127 (1989).
13. Fynbo, J. P. U. *et al.* No supernovae associated with two long-duration γ -ray bursts. *Nature*, doi:10.1038/nature05375 (this issue); preprint at (astro-ph/0608313) (2006).
14. Gal-Yam, A. *et al.* A novel explosive process is required for the γ -ray burst GRB 060614. *Nature*, doi:10.1038/nature05373 (this issue); preprint at (http://arxiv.org/astro-ph/0608257) (2006).
15. Della Valle, M. *et al.* An enigmatic long-lasting γ -ray burst not accompanied by a bright supernova. *Nature*, doi:10.1038/nature05374 (this issue); preprint at (http://arxiv.org/astro-ph/0608322) (2006).
16. Gehrels, N. *et al.* The Swift gamma ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
17. Price, P. A., Berger, E. & Fox, D. B. GRB 060614: redshift. *GCN Circ.* **5275** (2006).
18. Schaefer, B. E. & Xiao, L. GRB060614 is at high redshift, so no new class of gamma-ray bursts is required. Preprint at (http://arxiv.org/astro-ph/0608441) (2006).
19. Cobb, B. E. *et al.* Could GRB 060614 and its presumed host galaxy be a chance superposition? *Astrophys. J.* **651**, L85–L88 (2006).
20. Kaneko, Y. *et al.* Prompt and afterglow emission properties of gamma-ray bursts with spectroscopically identified supernovae. *Astrophys. J.* (in the press); preprint at (http://arxiv.org/astro-ph/0607110) (2006).
21. Norris, J. P. & Bonnell, J. T. Short gamma-ray bursts with extended emission. *Astrophys. J.* **643**, 266–275 (2006).
22. Mazets, E. P. *et al.* *Konus Catalog of Short GRBs* (http://www.ioffe.ru/LEA/shortGRBs/Catalog/). Preprint at (http://arxiv.org/astro-ph/astro-ph/0209219) (Ioffe Physico-Technical Institute, Laboratory for Experimental Astrophysics, 2002).
23. Norris, J. P. Implications of the lag-luminosity relationship for unified gamma-ray burst paradigms. *Astrophys. J.* **579**, 386–403 (2002).
24. Donaghy, T. Q. *et al.* HETE-2 localizations and observations of four short gamma-ray bursts: GRBs 010326B, 040802, 051211 and 060121. Preprint at (http://arxiv.org/astro-ph/0605570) (2006).
25. Zhang, W., Woosley, S. E. & MacFadyen, A. I. Numerical simulations of relativistic jets in collapsars. *Astrophys. J.* **586**, 356–371 (2003).
26. Woosley, S. E. & Bloom, J. S. The supernova–gamma-ray burst connection. *Annu. Rev. Astron. Astrophys.* **44**, 507–561 (2006).
27. Narayan, R., Piran, T. & Kumar, P. Accretion models of gamma-ray bursts. *Astrophys. J.* **557**, 949–957 (2001).
28. Faber, J. A. *et al.* Relativistic binary merger simulations and short gamma-ray bursts. *Astrophys. J.* **641**, L93–L96 (2006).
29. Golenetskii, S. *et al.* Konus-Wind observation of GRB 060614. *GCN Circ.* **5264** (2006).
30. Vanderspek, R. *et al.* Observations of the gamma ray burst GRB 030329: Evidence for a soft underlying component. *Astrophys. J.* **617**, 1251–1257 (2004).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.G. (gehrels@milkyway.gsfc.nasa.gov).

No supernovae associated with two long-duration γ -ray bursts

Johan P. U. Fynbo¹, Darach Watson¹, Christina C. Thöne¹, Jesper Sollerman^{1,3}, Joshua S. Bloom⁴, Tamara M. Davis¹, Jens Hjorth¹, Páll Jakobsson⁵, Uffe G. Jørgensen², John F. Graham⁶, Andrew S. Fruchter⁶, David Bersier⁷, Lisa Kewley⁸, Arnaud Cassan⁹, José María Castro Cerón¹, Suzanne Foley¹⁰, Javier Gorosabel¹¹, Tobias C. Hinse², Keith D. Horne¹², Brian L. Jensen¹, Sylvio Klose¹³, Daniel Kocevski⁴, Jean-Baptiste Marquette¹⁴, Daniel Perley⁴, Enrico Ramirez-Ruiz^{15,16}, Maximilian D. Stritzinger¹, Paul M. Vreeswijk^{17,18}, Ralph A. M. Wijers¹⁹, Kristian G. Woller², Dong Xu¹ & Marta Zub⁹

It is now accepted that long-duration γ -ray bursts (GRBs) are produced during the collapse of a massive star^{1,2}. The standard 'collapsar' model³ predicts that a broad-lined and luminous type Ic core-collapse supernova accompanies every long-duration GRB⁴. This association has been confirmed in observations of several nearby GRBs^{5–9}. Here we report that GRB 060505 (ref. 10) and GRB 060614 (ref. 11) were not accompanied by supernova emission down to limits hundreds of times fainter than the archetypal supernova SN 1998bw that accompanied GRB 980425, and fainter than any type Ic supernova ever observed¹². Multi-band observations of the early afterglows, as well as spectroscopy of the host galaxies, exclude the possibility of significant dust obscuration and show that the bursts originated in actively star-forming regions. The absence of a supernova to such deep limits is qualitatively different from all previous nearby long-duration GRBs and suggests a new phenomenological type of massive stellar death.

GRB 060505 and GRB 060614 were detected by the γ -ray Burst Alert Telescope onboard the dedicated GRB satellite Swift on 5.275 May 2006 and 14.530 June 2006, respectively^{11,12}. GRB 060505 was a faint burst with a duration of 4 s. GRB 060614 had a duration of 102 s and a pronounced hard-to-soft evolution. Both were rapidly localized by Swift's X-ray telescope. Subsequent follow-up of these bursts led to the discovery of their optical afterglows, locating them in galaxies at low redshift: GRB 060505 at $z = 0.089$ (ref. 13) and GRB 060614 at $z = 0.125$ (refs 14, 15). The relative proximity of these bursts engendered an expectation that a bright supernova would be discovered a few days after the bursts, as had been found just a few months before in another low-redshift Swift burst, GRB 060218 ($z = 0.033$)⁹, and in all previous well-observed nearby bursts^{1,5–8}.

We monitored the afterglows of GRB 060505 and GRB 060614 using a range of telescopes (see Supplementary Information for details). These led to early detections of the afterglows. We continued the monitoring campaign and obtained stringent upper limits on any re-brightening at the position of the optical afterglows up to 12 and 5 weeks after the bursts, respectively. The light curves obtained from

this monitoring are shown in Fig. 1. For GRB 060505 we detected the optical afterglow at a single epoch. All subsequent observations resulted in deep upper limits. For GRB 060614 we followed the decay of the optical afterglow in the R-band up to four nights after the burst. In later observations no source was detected to deep limits (see also refs 14, 15 for independent studies of this event). As seen in Fig. 1, the upper limits are far below the level seen in previous supernovae, and in particular previous supernovae associated with long-duration GRBs^{5–9}. For both GRBs our 3σ limits around the time of expected maximum of a supernova component are 80–100 times fainter than SN 1998bw would have appeared. The very deep limits for GRB 060505 from 23 May and 30 May places a 3σ upper limit more than 250 times fainter than SN 1998bw at a similar time. Hence, any associated supernova must have a peak magnitude in the R-band fainter than about -13.5 .

A concern in any attempt to uncover a supernova associated with a GRB is the presence of a poorly quantified level of extinction along the line of sight. In these cases, however, the levels of Galactic extinction in both directions are fortunately very low: $E(B-V) = 0.02$ (ref. 16). In the case of GRB 060505, our spatially resolved spectroscopy of the host galaxy allows us to use the Balmer emission line ratios to limit the dust obscuration at the location of the burst. The Balmer line ratio is consistent with no internal reddening. In the case of GRB 060614, the detection of the early afterglow in many bands, including the Swift ultraviolet bands¹⁷ UVW1 and UVW2, rules out significant obscuration of the source in the host galaxy and we conclude that there is no significant dust obscuration in either case (see also ref. 15).

Both GRBs were located in star-forming galaxies. The host galaxy of GRB 060505 has an absolute magnitude of about $M_B = -19.6$ and the spectrum displays the prominent emission lines typically seen in star-forming galaxies. The two-dimensional spectrum shows that the host galaxy emission seen at the position of the afterglow is due to a compact H II region in a spiral arm of the host (see the Supplementary Information for details). We estimate a star-formation rate of

¹Dark Cosmology Centre, ²Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark. ³Department of Astronomy, Stockholm University, AlbaNova, S-10691 Stockholm, Sweden. ⁴Department of Astronomy, University of California at Berkeley, 601 Campbell Hall, Berkeley, California 94720, USA. ⁵Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, UK. ⁶Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ⁷Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD, UK. ⁸University of Hawaii, Institute of Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ⁹Astronomisches Rechen-Institut (ARI), Zentrum für Astronomie der Universität Heidelberg (ZAH), Mönchhof Str. 12-14, 69120 Heidelberg, Germany. ¹⁰School of Physics, University College Dublin, Dublin 4, Ireland. ¹¹Instituto de Astrofísica de Andalucía (CSIC), Apartado de Correos, 3004, E-18080 Granada, Spain. ¹²SUPA Physics/Astronomy, University of St Andrews, St Andrews, KY16 9SS, Scotland, UK. ¹³Thüringer Landessternwarte Tautenburg, Sternwarte 5, 07778 Tautenburg, Germany. ¹⁴Institut d'Astrophysique de Paris, UMR70951 CNRS, Université Pierre et Marie Curie, 98 bis boulevard Arago, 75014 Paris, France. ¹⁵Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA. ¹⁶Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. ¹⁷European Southern Observatory, Alonso de Córdova 3107, Casilla 19001, Vitacura. ¹⁸Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile. ¹⁹Astronomical Institute 'Anton Pannekoek', Faculty of Science, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands.

$1M_{\odot} \text{ yr}^{-1}$ (where $M_{\odot}$ is the mass of the Sun) and a specific rate of about $4M_{\odot} \text{ yr}^{-1} (L/L^*)^{-1}$ (assuming an absolute magnitude corresponding to the B-band luminosity L^* of $M_B^* = -21$). The host galaxy of GRB 060614 is significantly fainter with an absolute magnitude of about $M_B = -15.3$. This is one of the least luminous GRB host gal-

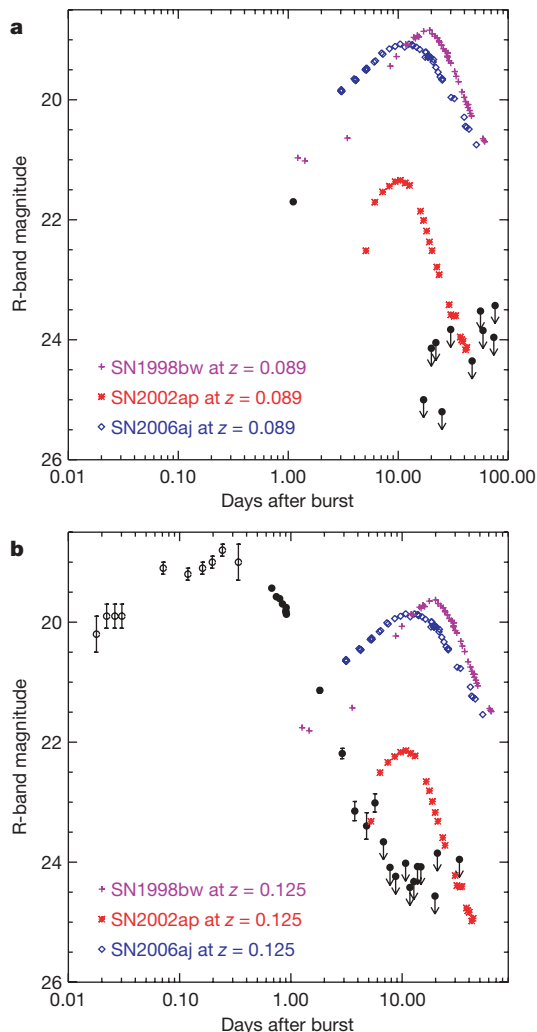


Figure 1 | No supernova associated with two nearby γ -ray bursts. Light curves of supernovae SN 1998bw, SN 2002ap and SN 2006aj, as they would have appeared at the redshift of GRB 060505 (a, top) and at the redshift of GRB 060614 (b, bottom). We also plot our afterglow detection in both cases (filled circles) and subsequent 3σ upper limits. We conclude that neither GRB 060505 nor GRB 060614 were associated with significant supernova emission down to very faint limits hundreds of times less luminous than the archetypal SN 1998bw. GRB 060505 was monitored with the Danish 1.5 m telescope, the Very Large Telescope and the Keck telescope over a period of 12 weeks, which allowed us to detect the afterglow on the first night after the burst and thereafter obtain very strict upper limits on associated supernova emission. GRB 060614 was observed with the Danish 1.5 m telescope almost every night up to 5 weeks after the burst, which enabled us to monitor the decay of the optical afterglow, and finally to detect the host galaxy. For GRB 060614 the early light curve is populated with data reported in the GRB Coordinates Network Circulars²⁹ (open circles). These supernova-less GRBs share no obvious characteristics in their prompt emission. GRB 060505 was one of the least luminous bursts discovered by Swift, with an isotropic-equivalent energy release of 1.2×10^{49} erg. It had a relatively short duration, and was single-peaked with a very faint afterglow. GRB 060614 was about a hundred times more luminous with an isotropic-equivalent energy release of 8.9×10^{50} erg, and it showed strong spectral evolution. Its optical afterglow brightened over the first day^{17,29}, reminiscent of GRB 970508³⁰. In the Supplementary Information we provide further details on the observations and data analysis. Error bars are 1σ .

axies ever detected, to our knowledge. We detect emission lines from hydrogen and oxygen and infer a star-formation rate of $0.014M_{\odot} \text{ yr}^{-1}$ (see Supplementary Information for details). The specific star-formation rate is $3M_{\odot} \text{ yr}^{-1} (L/L^*)^{-1}$. Sub- L^* and star-forming host galaxies like these two are ubiquitous among long-duration GRB host galaxies¹⁸. For comparison, the specific star-formation rates of the four previously studied nearby ($z < 0.2$) long-duration GRB host galaxies are 6, 7, 25 and $39M_{\odot} \text{ yr}^{-1} (L/L^*)^{-1}$ (refs 19, 20).

All spectroscopically confirmed supernova-GRBs have peak magnitudes within half a magnitude of SN 1998bw^{1,2}, and all photometrically identified supernova-GRBs have peak magnitudes within one and a half magnitudes of SN 1998bw^{1,21,22}. For X-ray flashes there has been evidence that associated supernovae span a somewhat wider range of luminosities^{9,23–25}, but still within the range of non-GRB-selected type Ic supernovae. The available data on all type Ic supernovae (including those unrelated to GRBs) show a distribution from roughly two magnitudes fainter in the V band than SN 1998bw to perhaps half a magnitude brighter^{1,2,12}. The faintest type Ic supernova known is SN 1997ef (classified as a broad-line supernova); but even a supernova as faint as this would easily have been detected in our observations of GRB 060505 and GRB 060614 (SN 2002ap shown in Fig. 1 had a luminosity similar to that of SN 1997ef). Any supernova associated with these two long-duration GRBs must therefore have been not only fainter than any supernova previously associated with a GRB or XRF but also substantially fainter than any non-GRB-related type Ic supernova seen until now.

The non-appearance of a supernova in these cases is a surprise and indicates that we have uncovered GRBs with quite different properties from those studied previously. It is possible that the origin of these bursts lies in one of the many supernova-less GRB progenitors suggested before the definitive association between GRBs and supernovae. Also, in a variant of the original collapsar model ‘fallback’-formed black holes or progenitors with relatively low angular momentum could produce supernova-less GRBs^{26–28}. Of the six long-duration GRBs or X-ray flashes known to be at low redshift ($z < 0.2$), two now have no associated supernova, so the fraction of supernova-less GRBs could be substantial.

Received 7 July; accepted 20 October 2006.

1. Zeh, A., Klose, S. & Hartmann, D. H. A systematic analysis of supernova light in gamma-ray burst afterglows. *Astrophys. J.* **609**, 952–961 (2004).
2. Woosley, S. E. & Bloom, J. S. The supernova–gamma-ray burst connection. *Annu. Rev. Astron. Astrophys.* **44**, 507–556 (2006).
3. Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
4. MacFadyen, A. I. & Woosley, S. E. Collapsars: Gamma-ray bursts and explosions in “failed supernovae”. *Astrophys. J.* **524**, 262–289 (1999).
5. Galama, T. J. et al. An unusual supernova in the error box of the gamma-ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
6. Hjorth, J. et al. A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
7. Stanek, K. Z. et al. Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
8. Malesani, D. et al., *Astrophys. J.* **609**, L5–L8 (2004).
9. Pian, E. et al. An optical supernova associated with X-ray flash 060218. *Nature* **442**, 1011–1013 (2006).
10. Hullinger, D. et al. GRB 060505 BAT refined analysis. *GCN Circ.* **5142** (2006).
11. Gehrels, N. et al. A new γ -ray burst classification scheme from GRB 060614. *Nature* doi:10.1038/nature05376 (this issue); preprint at (<http://arxiv.org/astro-ph/0610635>) (2006).
12. Richardson, D., Branch, D. & Baron, E. Absolute magnitude distributions and light curves of stripped-envelope supernovae. *Astron. J.* **131**, 2233–2244 (2006).
13. Ofek, E. O. et al. GRB 060505 – OT candidate + galaxy spectrum. *GCN Circ.* **5123** (2006).
14. Gal-Yam, A. et al. A novel explosive process is required for the γ -ray burst GRB 060614. *Nature* doi:10.1038/nature05373 (this issue); preprint at (<http://arxiv.org/astro-ph/0608257>) (2006).
15. Della Valle, M. et al. An enigmatic long-lasting γ -ray burst not accompanied by a bright supernova. *Nature* doi:10.1038/nature05374 (this issue); preprint at (<http://arxiv.org/astro-ph/0608322>) (2006).

16. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
17. Holland, S. T. GRB 060614: Swift/UVOT Observations. *GCN Circ.* **5255** (2006).
18. Fruchter, A. S. *et al.* Long gamma-ray bursts and core-collapse supernovae have different environments. *Nature* **441**, 463–468 (2006).
19. Sollerman, J. *et al.* On the nature of nearby GRB/SN host galaxies. *New Astron.* **11**, 103–115 (2005).
20. Sollerman, J. *et al.* Supernova 2006aj and the associated X-ray flash 060218. *Astron. Astrophys.* **454**, 503–509 (2006).
21. Soderberg, A. M. *et al.* An HST study of the supernovae accompanying GRB 040924 and GRB 041006. *Astrophys. J.* **636**, 391–399 (2006).
22. Ferrero, P. *et al.* The GRB 060218/SN 2006aj event in the context of other gamma-ray burst supernovae. *Astron. Astrophys.* **457**, 857–864 (2006).
23. Levan, A. *et al.* A deep search with the Hubble Space Telescope for late-time supernova signatures in the hosts of XRF 011030 and XRF 020427. *Astrophys. J.* **622**, 975–985 (2005).
24. Soderberg, A. M. *et al.* An HST search for supernovae accompanying X-ray flashes. *Astrophys. J.* **627**, 877–887 (2005).
25. Fynbo, J. P. U. *et al.* On the afterglow of the X-ray flash of 2003 July 23: photometric evidence for an off-axis gamma-ray burst with an associated supernova? *Astrophys. J.* **609**, 962–971 (2004).
26. Fryer, C. L., Young, P. A. & Hungerford, A. L. Explosive nucleosynthesis from GRB and hypernova progenitors: direct collapse versus fallback. *Astrophys. J.* **650**, 1028–1047 (2006).
27. Heger, A. *et al.* How massive single stars end their life. *Astrophys. J.* **591**, 288–300 (2003).
28. MacFadyen, A. I. in *AIP Conf. Proc. 662: Gamma-Ray Burst and Afterglow Astronomy 2001: A Workshop Celebrating the First Year of the HETE Mission (November 2001)* (eds Ricker, G. R. & Vanderspek, R. K.) 202–205 (American Institute of Physics, Melville, New York, 2003).
29. Schmidt, B. P. *et al.* SSO 1m observations. *GCN Circ.* **5258** (2006).
30. Pedersen, H. *et al.* Evidence for diverse optical emission from gamma-ray burst sources. *Astrophys. J.* **496**, 311–315 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge benefits from collaboration within the EU FP5 Research Training Network “Gamma-Ray Bursts: An Enigma and a Tool”. The Dark Cosmology Centre is funded by the DNRF. The observations presented here have been obtained from the ESO La Silla-Paranal observatory and from the Gemini Observatory. The Gemini Observatory is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (USA), the Particle Physics and Astronomy Research Council (UK), the National Research Council (Canada), CONICYT (Chile), the Australian Research Council (Australia), CNPq (Brazil), and CONICET (Argentina).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P.U.F. (jfyfynbo@dark-cosmology.dk).

LETTERS

An enigmatic long-lasting γ -ray burst not accompanied by a bright supernova

M. Della Valle¹, G. Chincarini^{2,3}, N. Panagia^{4,5,6}, G. Tagliaferri³, D. Malesani⁷, V. Testa⁸, D. Fugazza^{2,3}, S. Campana³, S. Covino³, V. Mangano⁹, L. A. Antonelli^{8,10}, P. D'Avanzo^{3,11}, K. Hurley¹², I. F. Mirabel¹³, L. J. Pellizza¹⁴, S. Piranomonte⁸ & L. Stella⁸

Gamma-ray bursts (GRBs) are short, intense flashes of soft γ -rays coming from the distant Universe. Long-duration GRBs (those lasting more than ~ 2 s) are believed to originate from the deaths of massive stars¹, mainly on the basis of a handful of solid associations between GRBs and supernovae^{2–7}. GRB 060614, one of the closest GRBs discovered, consisted of a 5-s hard spike followed by softer, brighter emission that lasted for ~ 100 s (refs 8, 9). Here we report deep optical observations of GRB 060614 showing no emerging supernova with absolute visual magnitude brighter than $M_V = -13.7$. Any supernova associated with GRB 060614 was therefore at least 100 times fainter, at optical wavelengths, than the other supernovae associated with GRBs¹⁰. This demonstrates that some long-lasting GRBs can either be associated with a very faint supernova or produced by different phenomena.

Following the discovery of the X-ray and optical afterglow of GRB 060614⁹, we observed it with the European Southern Observatory (ESO) 8.2 m Very Large Telescope (VLT). The spectrum of the host galaxy exhibits nebular emission lines (Fig. 1), which reveal ongoing star formation and the presence of young, massive stars. The specific star formation rate, normalized to the host luminosity (B-band magnitude $M_B = -15.5$), is about $2M_\odot \text{ yr}^{-1} L_*^{-1}$, where $M_\odot$ is the solar mass and L_* is the typical luminosity of field galaxies. This value is comparable to that exhibited by the Milky Way, and is at the low end of the distribution found for long-duration GRB hosts¹¹. The galaxy is also fainter than most GRB hosts¹².

Starting 15 h after the burst, we monitored the light curve of the optical transient associated with GRB 060614. Deep observations obtained in the R band (roughly corresponding to the V band in the GRB rest frame) up to 65 d after the burst did not reveal the emergence of a supernova component (Fig. 2; see also refs 13, 14); such a component has been often observed in a number of nearby long-duration GRBs. Our data are consistent with no supernova contribution (blue line). Adopting as a template the light curve of supernova SN 1998bw², we constrain the brightness of a supernova coincident with GRB 060614 to be at least 5.6 mag fainter than the template (3σ limit; green lines in Fig. 2). This corresponds to a peak absolute magnitude $M_V > -13.5$. Similar limiting magnitudes are obtained adopting different supernova light curve shapes. A brighter supernova (yellow lines in Fig. 2) would provide a totally inadequate fit. The faintness of a possible supernova at optical wavelengths was further confirmed by a series of about ten spectra obtained at the VLT

in the 4,500–8,000 Å wavelength range, between 2006 June 15 and 2006 July 30. None of them shows the broad undulations due to the very high expansion velocities ($\sim 30,000 \text{ km s}^{-1}$) typical of the supernovae¹⁵ associated with GRBs.

Such faintness cannot be due to dust extinction. First, the afterglow optical spectra are not particularly red (from our BVRIJK

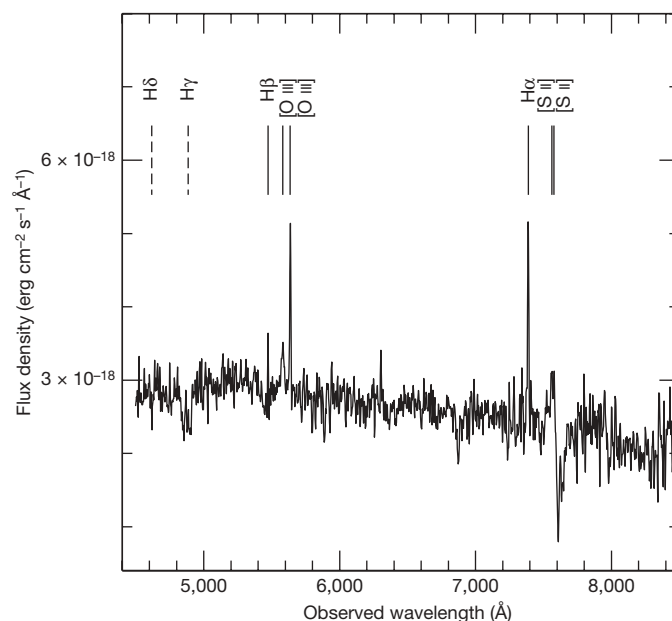


Figure 1 | Spectrum of the host galaxy of GRB 060614. This is the average of several observations taken with the VLT equipped with the FORS2 spectrograph in the period 2006 June 20 to July 30. From the emission features (marked with solid bars) we infer a redshift $z = 0.1254 \pm 0.0005$. This confirms the redshift proposed in ref. 28. $H\gamma$ and $H\delta$ are seen in absorption (dashed bars). The flux from the $H\alpha$ line, not corrected for internal extinction, amounts to $4.1 \times 10^{-17} \text{ erg cm}^{-2} \text{ s}^{-1}$ (corrected for slit loss). This corresponds to an unobscured star formation rate of $1.3 \times 10^{-2} M_\odot \text{ yr}^{-1}$. Given the faintness of the galaxy ($M_B \approx -15.5$), however, the specific star formation rate ($2M_\odot \text{ yr}^{-1} L_*^{-1}$, assuming for the absolute B-band magnitude of field galaxies $M_B^* = -21$) is not negligible. From the observed flux of the [O III] lines and the limits on [N II], we infer a metallicity larger than $\sim 1/20$ solar.

¹INAF, Osservatorio Astrofisico di Arcetri, largo E. Fermi 5, I-50125 Firenze, Italy. ²Dipartimento di Fisica, Università degli Studi di Milano-Bicocca, piazza delle Scienze 3, I-20126 Milano, Italy. ³INAF, Osservatorio Astronomico di Brera, via E. Bianchi 46, I-23807 Merate (Lc), Italy. ⁴Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ⁵Istituto Nazionale di Astrofisica, viale del Parco Mellini 84, I-00136, Roma, Italy. ⁶Supernova Ltd, Olde Yard Village #131, Northsound Road, Virgin Gorda, British Virgin Islands. ⁷International School for Advanced Studies (SISSA/ISAS), via Beirut 2–4, I-34014 Trieste, Italy. ⁸INAF, Osservatorio Astronomico di Roma, via di Frascati 33, I-00040 Monteporzio Catone (Roma), Italy. ⁹INAF, Istituto di Astrofisica Spaziale e Fisica Cosmica di Palermo, via U. La Malfa 153, I-90146 Palermo, Italy. ¹⁰ASI Science Data Center, via G. Galilei, I-00044 Frascati, Italy. ¹¹Dipartimento di Fisica e Matematica, Università dell'Insubria, via Valleggio 11, I-22100 Como, Italy. ¹²University of California, Berkeley, Space Sciences Laboratory, Berkeley, California 94720-7450, USA. ¹³European Southern Observatory, Alonso de Córdova 3107, Vitacura, Casilla 19001, Santiago 19, Chile. ¹⁴Instituto de Astronomía y Física del Espacio (CONICET/UBA), Casilla de Correos 67 Suc. 28, (1428) Buenos Aires, Argentina.

photometry we measure a spectral index $\beta = 0.94$ at ~ 1.7 d after the GRB). The afterglow is also bright in the ultraviolet⁹, where extinction would be more severe. X-ray spectra (V.M. *et al.*, manuscript in preparation) also show little absorbing material along the line of sight (yielding a rest-frame hydrogen column density $N_{\text{H}} < 2 \times 10^{20} \text{ cm}^{-2}$ at 90% confidence level). Furthermore, by modelling the broadband optical/X-ray spectral energy distribution (see Supplementary Fig. 1), we can estimate that the source was affected by less than 0.2 mag of extinction in the observed R band. Allowing for this amount of extinction, we can refine our limit for the supernova peak magnitude to $M_V > -13.7$.

So far only type-Ib/c events have been clearly associated with GRBs¹ and, had the progenitor of GRB 060614 been one of them, its expected colour at maximum light would be $B - V \approx 0.5 \pm 0.1$, as measured in well-observed events such as SN 2006aj¹⁰, SN 2002ap¹⁶ and SN 1998bw², as well in several other type-Ib/c supernovae. From the tables of ref. 17, we find that this colour corresponds to an effective temperature $T \approx 6,500 \text{ K}$, which, combined with the lower limit to the absolute magnitude, provides a strict upper limit to the bolometric luminosity, at maximum light, of $L < 10^{41} \text{ erg s}^{-1}$. Therefore the radius of the emitting region is constrained to be $R \approx \sqrt{L/(4\pi\sigma T^4)} < 2.8 \times 10^{14} \text{ cm}$ (where σ is the Stefan-Boltzmann constant). As the rise times to maximum light of type-Ib/c supernovae range between 10 and 20 d (refs 10, 18–20) in the V band, the upper limit to the radius implies an upper limit to the expansion velocity in the range 1,600–3,200 km s^{-1} . This value is

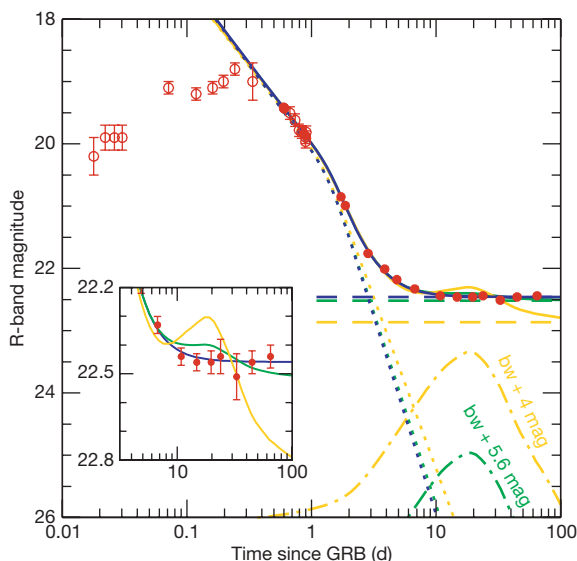


Figure 2 | R-band light curve of the GRB 060614 afterglow. Red open circles show data from the literature^{13,14,29} (not used in the fits); red filled circles represent our VLT data (see Supplementary Table 1). Error bars (smaller than symbols for most of our data) show the 1σ errors. Photometry was performed adopting large apertures in order to include all the flux from the host galaxy. Flux calibration was achieved by observing several Landolt standard fields. The data were modelled as the sum (solid lines) of three components: the afterglow (dotted lines), the host (dashed lines) and a supernova akin to SN 1998bw but rescaled in flux ('bw'; dot-dashed lines). The different colours correspond to different contributions from the supernova: no contribution (blue), a supernova fainter by 5.6 mag (green), and a supernova fainter by 4 mag (yellow). The model shown by green lines corresponds to the brightest supernova allowed by our data, $M_V > -13.5$, at the 3σ level. At the 2σ level, the limit is $M_V > -12.9$. The model shown by the yellow lines is clearly inadequate. The inset shows an expanded version of the plot around the peak of a putative supernova. The afterglow component is described by a broken power law with decay indices $\alpha_1 = 1.08 \pm 0.03$ and $\alpha_2 = 2.48 \pm 0.07$, respectively before and after the break at $t_{\text{break}} = 1.39 \pm 0.04 \text{ d}$ ($\chi^2/\text{d.o.f.} = 15.5/20$). Interpreting this as a jet break, the inferred jet half-opening angle is $\theta \approx 12^\circ$.

an order of magnitude smaller than that observed in supernovae associated with GRBs^{7,15}.

The low expansion velocity and the faint luminosity implied for a possible supernova progenitor are reminiscent of a class of very faint core-collapse supernovae recently discovered²¹ in the local Universe. They are of type II, have absolute magnitudes at maximum in the range $-13 > M_V > -15$, and show very small expansion velocities ($\sim 1,000 \text{ km s}^{-1}$). The properties of such objects may well be consistent with the available data on GRB 060614.

Faint type-II supernovae have been interpreted in terms of the collapse of massive stars with an explosion energy so small that most of the ^{56}Ni falls back onto the compact stellar remnant²². Such supernovae share properties with the present case, both in terms of observational characteristics and because they are expected to give rise to a black hole (which is believed to be necessary for the production of a GRB). However, the possibility that such supernova progenitors are able to produce GRBs has yet to be explored. In particular, the stellar envelope would need to be absent for the relativistic jets to emerge out of the star. GRB 060614 might thus be an example of a fallback supernova of type Ib/c. The small amount of nickel ($< 10^{-3} M_\odot$), responsible for the very low luminosity, might possibly also provide little heating to the ejecta, leading to an unusually low temperature ($T \approx 2,000 \text{ K}$) and allowing for larger expansion velocities (as $v \propto T^{-2}$). In any case, GRB 060614 may be the prototype of a new class of GRBs originating from a new kind of massive star death, different from those producing both classical long-duration (associated with bright type-Ib/c supernovae) and short-duration (possibly originating in binary system mergers²³) GRBs. Some evidence for this idea comes from the high-energy properties of this GRB, which contemporaneously exhibits features typical of both the long and short GRB classes⁸. Indeed, scenarios in which the GRB was not directly connected to a supernova explosion²⁴ cannot be excluded by our data (though they are not required). For example, our data would be compatible with a supernova exploding before the GRB²⁵. Also, a binary merger mechanism²⁶, similar to that proposed to power short-duration GRBs, or some type of collapsar model²⁷, are not expected to produce a supernova. These results challenge the commonly accepted scenario in which long-duration GRBs are produced only together with very bright supernova explosions. Not all GRBs are produced in such a way.

Received 9 August; accepted 25 October 2006.

1. Woosley, S. E. & Bloom, J. S. The supernova-gamma-ray burst connection. *Annu. Rev. Astron. Astrophys.* **44**, 507–556 (2006).
2. Galama, T. J. *et al.* An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
3. Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
4. Hjorth, J. *et al.* A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
5. Malesani, D. *et al.* SN 2003lw and GRB 031203: A bright supernova for a faint gamma-ray burst. *Astrophys. J.* **609**, L5–L8 (2004).
6. Pian, E. *et al.* An optical supernova associated with the X-ray flash XRF 060218. *Nature* **442**, 1011–1013 (2006).
7. Campana, S. *et al.* The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* **442**, 1008–1010 (2006).
8. Gehrels, N. *et al.* A new γ -ray burst classification scheme from GRB 060614. *Nature* doi:10.1038/nature05376 (this issue).
9. Parsons, A. M. *et al.* GRB 060614: Swift detection of a burst with a bright optical and X-ray counterpart. *GCN Circ.* 5252 (2006).
10. Ferrero, P. *et al.* The GRB 060218/SN 2006aj event in the context of other gamma-ray burst supernovae. *Astron. Astrophys.* **457**, 857–864 (2006).
11. Christensen, L., Hjorth, J. & Gorosabel, J. UV star-formation rates of GRB host galaxies. *Astron. Astrophys.* **425**, 913–926 (2004).
12. Fruchter, A. S. *et al.* Long γ -ray bursts and core-collapse supernovae have different environments. *Nature* **441**, 463–468 (2006).
13. Gal-Yam, A. *et al.* A novel explosive process is required for the γ -ray burst GRB 060614. *Nature* doi:10.1038/nature05373 (this issue).
14. Fynbo, J. P. U. *et al.* No supernovae associated with two long-lasting γ -ray bursts. *Nature* doi:10.1038/nature05375 (this issue).
15. Patat, F. *et al.* The metamorphosis of SN 1998bw. *Astrophys. J.* **555**, 900–917 (2001).

16. Foley, R. *et al.* Optical photometry and spectroscopy of the SN 1998bw-like type Ic supernova 2002ap. *Publ. Astron. Soc. Pacif.* **115**, 1220–1235 (2003).
17. Romaniello, M., Panagia, N., Scuderi, S. & Kirshner, R. P. Accurate stellar population studies from multiband photometric observations. *Astron. J.* **123**, 915–940 (2002).
18. Della Valle, M. *et al.* Hypernova signatures in the late rebrightening of GRB 050525A. *Astrophys. J.* **642**, L103–L106 (2006).
19. Hamuy, M. Observed and physical properties of core-collapse supernovae. *Astrophys. J.* **582**, 905–914 (2003).
20. Panagia, N. in *Supernovae and Gamma-Ray Bursters* (ed. Weiler, K. W.) 113–144 (Lecture Notes in Physics, Vol. 598, Springer, Berlin, 2003).
21. Pastorello, A. *et al.* Low luminosity type II supernovae: Spectroscopic and photometric evolution. *Mon. Not. R. Astron. Soc.* **347**, 74–94 (2004).
22. Nomoto, K. *et al.* in *Stellar Collapse* (ed. Fryer, C.) 277–325 (Astrophysics and Space Science Library, Vol. 302, Kluwer Academic, Dordrecht, 2004).
23. Eichler, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrino bursts and γ -rays from coalescing neutron stars. *Nature* **340**, 126–128 (1989).
24. Blinnikov, S. I. & Postnov, K. A. A mini-SN model for optical afterglow of GRBs. *Mon. Not. R. Astron. Soc.* **293**, L29–L32 (1998).
25. Vietri, M. & Stella, L. A. Gamma-ray burst model with small baryon contamination. *Astrophys. J.* **507**, L45–L48 (1998).
26. Belczynski, K., Bulik, T. & Rudak, B. Study of gamma-ray burst binary progenitors. *Astrophys. J.* **571**, 394–412 (2002).
27. Woosley, S. E., Zhang, W. & Heger, A. in *Gamma-Ray Burst and Afterglow Astronomy 2001: A Workshop Celebrating the First Year of the HETE Mission* (eds Ricker, G. R. & Vanderspek, R.) 185–192 (AIP Conf. Proc., Vol. 662, American Institute of Physics, New York, 2003).
28. Price, P. A., Berger, E. & Fox, D. B. GRB 060614: Redshift. *GCN Circ.* **5275** (2006).
29. French, F., Melady, G., Hanlon, L., Jelinek, M. & Kubánek, P. GRB060614: Watcher observation. *GCN Circ.* **5257** (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is based on data collected at the Very Large Telescope operated by the European Southern Observatory. We acknowledge support from the observing staff. This research is supported in Italy by ASI, MIUR-PRIN and INAF-PRIN grants.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.D.V. (massimo@arcetri.astro.it).

A novel explosive process is required for the γ -ray burst GRB 060614

A. Gal-Yam¹, D. B. Fox², P. A. Price³, E. O. Ofek¹, M. R. Davis⁴, D. C. Leonard⁴, A. M. Soderberg¹, B. P. Schmidt⁵, K. M. Lewis⁵, B. A. Peterson⁵, S. R. Kulkarni¹, E. Berger^{6,7}, S. B. Cenko¹, R. Sari¹, K. Sharon⁸, D. Frail⁹, D.-S. Moon¹, P. J. Brown², A. Cucchiara², F. Harrison¹, T. Piran¹⁰, S. E. Persson^{6,7}, P. J. McCarthy^{6,7}, B. E. Penprase¹¹, R. A. Chevalier¹² & A. I. MacFadyen^{13,14}

Over the past decade, our physical understanding of γ -ray bursts (GRBs) has progressed rapidly, thanks to the discovery and observation of their long-lived afterglow emission. Long-duration (≥ 2 s) GRBs are associated with the explosive deaths of massive stars ('collapsars', ref. 1), which produce accompanying supernovae^{2–5}; the short-duration ($\lesssim 2$ s) GRBs have a different origin, which has been argued to be the merger of two compact objects^{6–9}. Here we report optical observations of GRB 060614 (duration ~ 100 s, ref. 10) that rule out the presence of an associated supernova. This would seem to require a new explosive process: either a massive collapsar that powers a GRB without any associated supernova, or a new type of 'engine', as long-lived as the collapsar but without a massive star. We also show that the properties of the host galaxy (redshift $z = 0.125$) distinguish it from other long-duration GRB hosts and suggest that an entirely new type of GRB progenitor may be required.

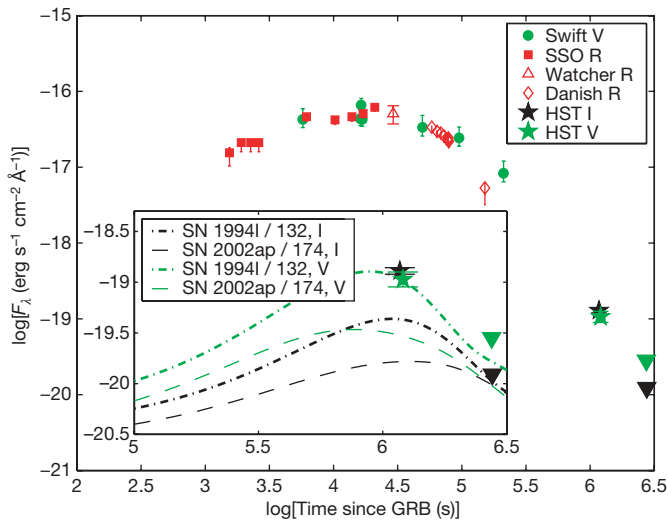
On 14 June 2006, at 12:43 UT, the burst alert telescope (BAT) on board the Swift satellite detected GRB 060614¹⁰, with a duration of 102 s. Detailed information was collected by the Swift BAT, X-ray telescope (XRT) and ultraviolet-optical telescope¹¹ (UVOT). In particular, the burst showed strong variability during much of that period, as confirmed by parallel observations by the Konus-Wind satellite¹², indicating sustained energy injection from an active engine, rather than the early onset of the afterglow (radiation from the interaction of an expanding outflow). We began observing this event ~ 26 min later using the 40-inch telescope at Siding Springs Observatory (SSO). The evolution of the optical radiation from this event as traced by our data, augmented by Swift observations and additional data from the literature, is shown in Fig. 1 (see Supplementary Information for data tables). As the optical source decayed, we noticed that it was apparently superposed on a faint dwarf host galaxy. On 19 June 2006 UT we obtained a spectrum of the host using the GMOS-S spectrograph mounted on the Gemini South 8-m telescope at Cerro Pachon, Chile. From this spectrum we derived the redshift of the host galaxy, $z = 0.125$, which is a low value for long GRBs. We confirmed this redshift with a higher quality spectrum obtained using the same instrument on 15 July 2006 UT (Supplementary Information section 2). Previous long GRBs at such low redshifts showed clear signatures of the underlying supernova

explosions at comparable age post-burst^{3,13}. However, such signatures were lacking in the case of this long GRB¹⁴.

Thus motivated, we undertook target-of-opportunity observations with the Hubble Space Telescope (HST). We observed the location of GRB 060614 using the Wide Field and Planetary Camera 2 (WFPC2) on board HST on 27–28 June 2006 UT, and again using the Advanced Camera for Surveys (ACS) on 15–16 July, 29 July and 8 September 2006 UT. Inspection of the data (Fig. 2) reveals a point source offset from the GRB host nucleus, which is well-detected in our first-epoch WFPC2 observations, and is apparently gone during our next visit. We identify this object as the optical afterglow of GRB 060614, and derive its brightness using image-subtraction methods. These high-resolution HST data strongly support the association of the GRB with the $z = 0.125$ host (Supplementary Information section 1). Our analysis (Fig. 1) shows that our HST detection is probably dominated by the afterglow (that is, residual decaying radiation from the interaction of the GRB ejecta with itself and/or the surrounding material), rather than a possible supernova (whose optical radiation is dominated by energy released from radioactive decay of newly synthesized elements, mostly ⁵⁶Ni), which is not required by the data. Any putative supernova component must be more than 100 times fainter than the faintest event previously known to be associated with a long GRB (supernova SN 2006aj/GRB 060218^{13,15}; Fig. 1). In fact, such a supernova (absolute V-band magnitude $M_V > -12.3$ mag, assuming V-band extinction $A_V < 0.2$; ref. 16) would be fainter than any supernova ever observed¹⁷. A conservative upper limit on the amount of synthesized ⁵⁶Ni is $5 \times 10^{-4} M_\odot$ (assuming that supernova peak luminosity scales with Ni mass¹⁸), which is more than two orders of magnitude less than the typical amount synthesized by long GRB/supernovae. Our HST data thus indicate that this GRB was not associated with a radioactively-powered event similar to any known supernova.

Furthermore, our HST and ground-based data reveal that the properties of the host of GRB 060614 and its environment are unusual when compared to those of the large sample of previously observed long GRBs. In particular, the star formation rate that we measure from the spectrum of the host, $0.0084 M_\odot \text{ yr}^{-1}$, is very small, and even the specific star formation rate, correcting for the low luminosity of this dwarf galaxy ($M_B = -15.9$ mag) is about ten times

¹Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA. ²Department of Astronomy and Astrophysics, Pennsylvania State University, 525 Davey Lab, University Park, Pennsylvania 16802, USA. ³Institute of Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822-1897, USA. ⁴Department of Astronomy, San Diego State University, San Diego, California 92182, USA. ⁵Research School of Astronomy and Astrophysics, Australian National University, Mt Stromlo Observatory, via Cotter Road, Weston Creek, Australian Capital Territory 2611, Australia. ⁶Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, California 91101, USA. ⁷Princeton University Observatory, Peyton Hall, Ivy Lane, Princeton, New Jersey 08544, USA. ⁸Department of Astrophysics, Tel Aviv University, 69978 Tel Aviv, Israel. ⁹National Radio Astronomy Observatory, PO Box 0, Socorro, New Mexico 87801, USA. ¹⁰Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel. ¹¹Department of Physics & Astronomy, Pomona College, 610 N. College Avenue, Claremont, California 91711, USA. ¹²Department of Astronomy, University of Virginia, PO Box 400325, Charlottesville, Virginia 22904, USA. ¹³Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA. ¹⁴Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, New York 10003, USA.



below that of typical long GRBs¹⁹. Moreover, the location of the GRB, well-separated from areas of the host which are brightest at rest-frame ultraviolet wavelengths, is atypical for a long GRB (only 1–2 bursts out of 32 (ref. 20) lie in locations that are fainter in the ultraviolet; see Supplementary Information section 2).

Considering the entire set of observations available for this event, the emerging picture is a puzzling one. On the one hand, the high-energy (γ -ray) duration of this burst is only consistent with that of a long GRB. On the other hand, the lack of an associated supernova is inconsistent with an origin in a massive, rapidly-rotating star undergoing a core-collapse supernova explosion (a collapsar¹)—the popular, observationally supported model for long GRBs. Furthermore, the environment and host galaxy properties of this event stand out from among those of numerous other long GRBs observed so far^{19,20}.

One may conjecture that all long GRBs result from massive collapsars, but that most of these are associated with supernovae with a

Figure 1 | Temporal evolution of the optical transient associated with GRB 060614. The main figure shows Swift V-band observations from UVOT (green filled circles), our SSO R-band data (red filled squares), augmented by R-band data from the Danish¹⁴ and Watcher²¹ groups (open red triangles and diamonds), along with our late time HST V- and I-band detections (green and black stars) and upper limits (green and black filled inverted triangles; see Fig. 2). The contribution of the host galaxy, estimated as detailed in Supplementary Information section 2, was removed from the photometry. Error bars (1σ) include root-mean-square photometric errors and calibration uncertainties added in quadrature, and are sometimes smaller than the symbols. The inset shows a comparison of our HST detections and upper limits with scaled-down light curves of supernovae, properly k-corrected and time-dilated¹⁸. The brightest allowed supernova is obtained by assuming the steeply declining light curve of SN 1994I, scaled down by a factor of 132 (heavy green (V) and black (I) dash-dot lines) and the maximal amount of extinction allowed by the analysis of early UVOT and XRT data¹⁶. In this case the absolute peak magnitude of the supernova will be $M_V = -12.3$, fainter than any supernova ever detected in the nearby Universe, synthesizing only $\sim 5 \times 10^{-4} M_\odot$ of ^{56}Ni (assuming that Ni mass scales with peak luminosity¹⁸). More likely scenarios involving supernovae with more slowly-decaying light curves—such as SN 2002ap^{25,26}, scaled down by a factor of 174 (thin green (V) and black (I) dashed lines), which is similar to the faintest GRB-associated supernova SN 2006aj—would impose more stringent limits on the luminosity and ^{56}Ni production of a putative supernova (by factors of ~ 2). At least the I-band emission detected during our first HST visit must be dominated by the GRB afterglow emission, with but a fraction of the light coming from a peaking faint supernova. Note that the data are well-explained without invoking any supernova-like component, with the optical afterglow roughly following the late X-ray decline rate¹¹ (to which a supernova is not expected to contribute).

range of properties (as observed so far^{18,21}) and a minority (the first unambiguous example of which is GRB 060614¹⁴) synthesize very little ^{56}Ni , and do not produce an optically luminous supernova. If this is indeed the case, the putative massive progenitor star is probably different from those of long GRBs, given its remarkable environment and host.

Alternatively, in view of the host galaxy properties, one might suggest that this event may not be associated with a massive stellar

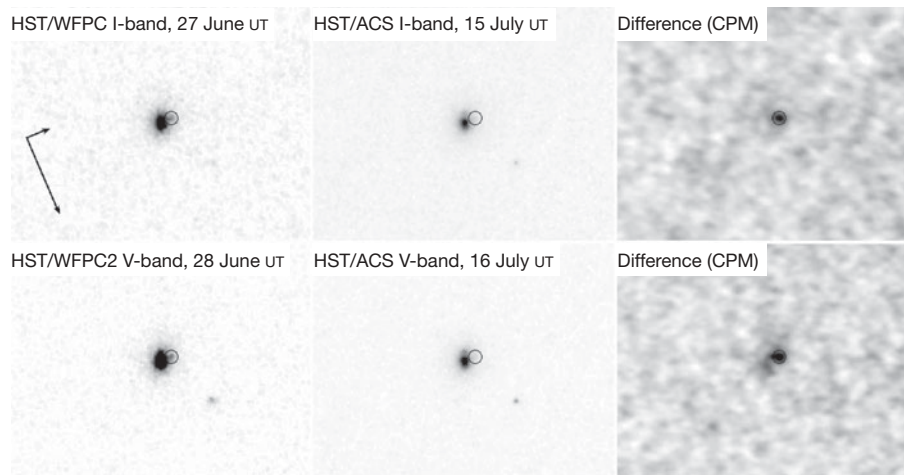


Figure 2 | HST observations of the location of GRB 060614. Top panels, I-band (F814W filter) data; bottom panels, V-band (F606W) data. Images were obtained using the WFPC2 camera (leftmost panels, 6,000 s total exposure time; UT dates as marked) and the ACS camera (middle panels, 3,600 s). Rightmost panels show the difference between the first epoch images and a third epoch visit with ACS (I-band; 29 July 2006 UT, total exposure time 4,840 s) or the second epoch in the case of the V-band, calculated using the image subtraction method CPM²⁷. WFPC2 data were reduced using custom scripts²⁸ and ACS data were reduced and photometry carried out in the standard manner with IRAF/multidrizze and then calibrated to standard bands²⁹. We used calibrated, nearby, isolated, compact sources to establish a calibration grid of Johnson-V and Cousins-I local standards and photometered the afterglow with respect to this grid

using the image-subtraction-based photometry pipeline mkdiffrc²⁷. A similar comparison between the second and third I-band (16 and 29 July UT) and V-band (15 July and 8 September) HST visits shows no residual to a 4σ upper limit of $I(V) = 27.5(27.75)$ mag, indicating that the optical transient was undetectable during our second visit. The resulting photometry and upper limits are reported in Fig. 1 and Supplementary Information section 3. Note the overall regular structure of this faint dwarf host and the peripheral location of the optical transient (see Supplementary Information section 2 for further details). The orientation of the images is marked with a long arrow due north and a short arrow due east. The length of the long arrow is $2.5''$ for scale. The black circle in all panels indicates the location of the afterglow as derived from the subtraction frames (right panels).

progenitor. Such a lower-mass long-lived progenitor system may be similar to those of short GRBs, which have been shown to reside in host galaxies of all types, including elliptical galaxies with virtually no young, massive stars^{6,7,8}, and in the outskirts of dwarf galaxies²², indicating a mature (rather than short-lived) population of progenitors (lacking associated supernovae). If this is the case, we predict that eventually a long event like GRB 060614 will be discovered in an elliptical host. If GRB 060614 and short GRBs share a common physical origin, maintaining its proposed identification as a compact binary merger is a challenge to the current consensus view, according to which this process lasts for but a fraction of a second²³.

Finally, GRB 060614 may be the first example of a new class of GRBs, different from both typical long events (which are associated with supernovae and powered by infall onto a newly formed black hole) and short events (which may come from compact binary mergers). Regardless of the ultimate resolution of this puzzle, it is already obvious that the elegant simple picture—consisting of two groups of GRBs with distinct physical origins (long GRBs from supernova/collapsars and short GRBs from binary mergers), which was briefly consistent with GRB observations and theory—must now be revised.

Received 11 August; accepted 20 October.

1. MacFadyen, A. I., Woosley, S. E. & Heger, A. Supernovae, jets, and collapsars. *Astrophys. J.* **550**, 410–425 (2001).
2. Galama, T. J. *et al.* An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
3. Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
4. Hjorth, J. *et al.* A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
5. Malesani, D. *et al.* SN 2003lw and GRB 031203: A bright supernova for a faint gamma-ray burst. *Astrophys. J.* **609**, L5–L8 (2004).
6. Gehrels, N. *et al.* A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$. *Nature* **437**, 851–854 (2005).
7. Bloom, J. S. *et al.* Closing in on a short-hard burst progenitor: Constraints from early-time optical imaging and spectroscopy of a possible host galaxy of GRB 050509b. *Astrophys. J.* **638**, 354–368 (2006).
8. Berger, E. *et al.* The afterglow and elliptical host galaxy of the short γ -ray burst GRB 050724. *Nature* **438**, 988–990 (2005).
9. Hjorth, J. *et al.* The optical afterglow of the short γ -ray burst GRB 050709. *Nature* **437**, 859–861 (2005).
10. Parsons, A. M. *et al.* GRB 060614: Swift detection of a burst with a bright optical and X-ray counterpart. *GCN Circ.* **5252** (2006).
11. Gehrels, N. *et al.* A new γ -ray burst classification scheme from GRB 060614. *Nature* doi:10.1038/nature05376 (this issue).
12. Golenetskii, S. *et al.* Konus-wind observation of GRB 060614. *GCN Circ.* **5264** (2006).
13. Pian, E. *et al.* An optical supernova associated with the X-ray flash XRF 060218. *Nature* **442**, 1011–1013 (2006).
14. Fynbo, J. P. U. *et al.* No supernovae associated with two long-duration γ -ray bursts. *Nature* doi:10.1038/nature05375 (this issue).
15. Modjaz, M. *et al.* Early-time photometry and spectroscopy of the fast evolving SN 2006aj associated with GRB 060218. *Astrophys. J.* **645**, L21–L24 (2006).
16. Della Valle, M. *et al.* An enigmatic long-lasting γ -ray burst not accompanied by a bright supernova. *Nature* doi:10.1038/nature05374 (this issue).
17. Pastorello, A. *et al.* Low-luminosity type II supernovae: spectroscopic and photometric evolution. *Mon. Not. R. Astron. Soc.* **347**, 74–94 (2004).
18. Soderberg, A. M. *et al.* An HST study of the supernovae accompanying GRB 040924 and GRB 041006. *Astrophys. J.* **636**, 391–399 (2006).
19. Christensen, L., Hjorth, J. & Gorosabel, J. UV star-formation rates of GRB host galaxies. *Astron. Astrophys.* **425**, 913–926 (2004).
20. Fruchter, A. S. *et al.* Long γ -ray bursts and core-collapse supernovae have different environments. *Nature* **441**, 463–468 (2006).
21. Zeh, A., Klose, S. & Hartmann, D. H. A systematic analysis of supernova light in gamma-ray burst afterglows. *Astrophys. J.* **609**, 952–961 (2004).
22. Fox, D. B. *et al.* The afterglow of GRB 050709 and the nature of the short-hard γ -ray bursts. *Nature* **437**, 845–850 (2005).
23. Narayan, R., Piran, T. & Kumar, P. Accretion models of gamma-ray bursts. *Astrophys. J.* **557**, 949–957 (2001).
24. French, J., Melady, G., Hanlon, L., Jelinek, M. & Kubanek, P. GRB060614: Watcher observation. *GCN Circ.* **5257** (2006).
25. Gal-Yam, A., Ofek, E. O. & Shemmer, O. Supernova 2002ap: the first month. *Mon. Not. R. Astron. Soc.* **332**, L73–L77 (2002).
26. Foley, R. J. *et al.* Optical photometry and spectroscopy of the SN 1998bw-like type Ic supernova 2002ap. *Publ. Astron. Soc. Pacif.* **115**, 1220–1235 (2003).
27. Gal-Yam, A. *et al.* On the progenitor of SN 2005gl and the nature of type II_n supernovae. *Astrophys. J.* (in the press); preprint at (<http://arxiv.org/astro-ph/0608029>) (2006).
28. Gal-Yam, A., Maoz, D. & Sharon, K. Supernovae in deep Hubble Space Telescope galaxy cluster fields: cluster rates and field counts. *Mon. Not. R. Astron. Soc.* **332**, 37–48 (2002).
29. Ofek, E. O. *et al.* GRB 060505: A possible short-duration gamma-ray burst in a star-forming region at a redshift of $z = 0.09$. *Astrophys. J.* (submitted).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.G. and E.B. acknowledge support by NASA through Hubble Fellowships. S.R.K. is supported by NSF and NASA. The HST campaign combined resources from our approved programmes (principal investigators S.R.K. and D.B.F.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.G. (avishay@astro.caltech.edu).

LETTERS

A deep dynamo generating Mercury's magnetic field

Ulrich R. Christensen¹

Mercury has a global magnetic field of internal origin and it is thought that a dynamo operating in the fluid part of Mercury's large iron core is the most probable cause. However, the low intensity of Mercury's magnetic field—about 1% the strength of the Earth's field—cannot be reconciled with an Earth-like dynamo. With the common assumption that Coriolis and Lorentz forces balance in planetary dynamos¹, a field thirty times stronger is expected. Here I present a numerical model of a dynamo driven by thermo-compositional convection associated with inner core solidification. The thermal gradient at the core–mantle boundary is subadiabatic^{2,3}, and hence the outer region of the liquid core is stably stratified with the dynamo operating only at depth, where a strong field is generated. Because of the planet's slow rotation the resulting magnetic field is dominated by small-scale components that fluctuate rapidly with time. The dynamo field diffuses through the stable conducting region, where rapidly varying parts are strongly attenuated by the skin effect, while the slowly varying dipole and quadrupole components pass to some degree. The model explains the observed structure and strength of Mercury's surface magnetic field and makes predictions that are testable with space missions both presently flying and planned.

During its fly-bys in 1974/75, the Mariner 10 spacecraft recorded an intrinsic magnetic field with mean surface intensity of about 450 nT. The limited data do not allow for an accurate determination of its geometry, but it seems clear that the field has a substantial dipole component nearly aligned with the rotation axis^{4,5}. The contribution of the axial quadrupole cannot be constrained and may be rather large⁵. A possible origin by remanent crustal magnetization is difficult to reconcile with the large-scale field geometry, although ways have been suggested for how this could arise⁶. Still, a self-sustained dynamo in the iron core is the most plausible cause.

The field strength B inside a planetary dynamo is expected to be such that the Lorentz force balances the Coriolis force. Frequently it is assumed that this is equivalent to the Elsasser number $A = B^2/(\mu_0 \lambda \rho \Omega)$ being of the order of one¹ (here λ is magnetic diffusivity, ρ is density and Ω is rotation rate), although the actual range in dynamos may be $A \approx 0.1$ –100 (ref. 7). However, extrapolating Mercury's observed field to the core–mantle boundary implies $A \approx 10^{-4}$ at the top of the core. The dilemma would be solved if Mercury possesses a strong field ($A \approx 1$) inside the dynamo region, but only a small fraction can escape through the core surface. For example, the toroidal part of the magnetic field (which has no radial component and is confined to the core) might be much stronger than the poloidal field (whose field lines do not need to close inside the core). In most geodynamo models both components have similar strength, but some models with a very small solid inner core⁸ or a large inner core⁹ exhibit a low poloidal-to-toroidal ratio. In another model with a large inner core¹⁰ the magnetic field inside the dynamo is dominated by higher-order components, whose amplitudes decrease rapidly with radius outside the core. However, in all these models the surface field is still too strong by a factor of ~ 10 or is dominated by non-axisymmetric higher-multipole components.

While the latter cannot be excluded for Mercury because of limited data, there is no positive evidence for it either.

Thermal evolution models^{2,3} predict a heat flow at Mercury's core–mantle boundary of several mW m^{-2} . This is only a fraction of the heat that can be conducted along the core adiabat ($>10 \text{ mW m}^{-2}$), implying a thermally stable outer core. The evolution models also suggest that Mercury has a solid inner core, whose radius has a wide possible range, depending mainly on the unknown sulphur concentration. Hence compositional convection, driven by rejection of sulphur from the growing inner core in the cooling planet, could power the dynamo. For a full treatment of such a system the evolution of stratification by double-diffusive convection¹¹ must be studied. In a simplified model, the combined effects of buoyancy by the deviation of the light-element concentration χ' from its mean in the fluid core and by the temperature deviation T' from an adiabatic reference state can be described by a single variable, the co-density¹² $C = \rho \alpha T' + \Delta \rho \chi'$, where $\Delta \rho$ is compositional density contrast and α is thermal expansivity. At the core–mantle boundary, where the compositional flux is zero, the co-density flux (or buoyancy flux) is negative. The positive flux of χ' at the inner core boundary (ICB) is controlled by the cooling rate and can be augmented by a locally superadiabatic heat flow¹³ resulting from the latent heat of solidification L . To give an example, with a core–mantle boundary heat flow of 5 mW m^{-2} , the integrated co-density flux Q_B is $6,000 \text{ kg s}^{-1}$ at the ICB and $-22,500 \text{ kg s}^{-1}$ at the core–mantle boundary (following the simplified model of ref. 13 and assuming an inner-core radius of $r_i = 648 \text{ km}$, core radius $r_o = 1,850 \text{ km}$, $\Delta \rho = 80 \text{ kg m}^{-3}$, melting temperature gradient $2 \times 10^{-8} \text{ K Pa}^{-1}$, thermal conductivity 43 W m K^{-1} , $L = 3.2 \times 10^5 \text{ J kg}^{-1}$, $\alpha = 3 \times 10^{-5} \text{ K}^{-1}$, $\rho = 8,200 \text{ kg m}^{-3}$, $T = 2,200 \text{ K}$, and specific heat $c_p = 700 \text{ J kg}^{-1} \text{ K}^{-1}$). The different signs of the boundary fluxes mean that the fluid core is stably stratified in its upper part and unstable at depth.

A numerical dynamo model is used to study magnetic field generation in a partly stable rotating fluid shell. Its set-up is similar to previous models^{7–10}, but the transport equation for temperature is replaced by a formally identical one for co-density¹⁴. We specify a homogeneous flux of C at the ICB and $C = 0$ on the core–mantle boundary. A homogeneous sink term ε inside the fluid shell represents the mixing of sulphur into the core fluid and the generation of negative buoyancy by subadiabaticity. Dynamo models cannot be calculated for planetary values of all control parameters, but it is possible to match some important properties. An essential one is the magnetic Reynolds number, $\text{Rm} = U/\lambda D$, with U the characteristic velocity and shell thickness $D = r_o - r_i$. A scaling law derived from previous dynamo simulations⁷ predicts $\text{Rm} \propto E_\lambda^{-1} \text{Ra}_Q^{2/5}$, where $E_\lambda = \lambda/(\Omega D^2)$ is the magnetic Ekman number and $\text{Ra}_Q^* = g_o Q_B/(4\pi r_o r_i D^2 \rho \Omega^3)$ is a modified Rayleigh number that is based here on the integrated co-density flux Q_B at the ICB (g_o is gravity at r_o). For Mercury, $E_\lambda \approx (0.5\text{--}1) \times 10^{-6}$ and $\text{Ra}_Q^* \approx (3\text{--}12) \times 10^{-8}$, assuming $r_i/r_o = 0.35\text{--}0.5$ and a plausible range for Q_B . Model parameters are $(E_\lambda, \text{Ra}_Q^*) = (10^{-5}, 1.08 \times 10^{-5})$ in case I and $(3.3 \times 10^{-5}, 2 \times 10^{-4})$ in case II. The combination $E_\lambda^{-1} \text{Ra}_Q^{2/5}$ is

¹Max-Planck Institute for Solar System Research, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany.

of the order of 1,000 for both models and for Mercury; hence the model and planetary values of Rm should agree.

Previous models^{7,15} suggest that a local Rossby number, $Ro_\ell = U/\Omega\ell \propto Q_B^{1/2} D^{-4/3} \Omega^{-7/6}$, controls whether the axial dipole or higher multipoles dominate the magnetic field (ℓ is a flow length

scale). When Ro_ℓ exceeds a threshold that depends on various factors but is probably always less than one, the field is non-dipolar. The scaling law puts Mercury with $Ro_\ell \approx 10$ into the non-dipolar regime¹⁵, mainly because its rotation is 60 times slower than Earth's, in apparent contradiction to the observation of a dipolar field. $Ro_\ell = 10$ cannot be reached, but model parameters are such that the dynamo field is multipolar.

In case I, with $r_i/r_o = 0.35$, the convective circulation is restricted to the unstable layer in the equatorial plane (Fig. 1a). It extends into the stable region in the form of columns aligned with the rotation axis (Fig. 1b), similar to what has been found in stratified rotating convection¹⁶. The (dynamic) co-density stratification is unstable in the lower 40% of the fluid core (Fig. 1d). In the unstable region, toroidal and poloidal flow components (the latter involves radial motion) are of similar size (Fig. 1e), but in the stable region the flow is mainly toroidal (purely horizontal) and to a substantial degree axisymmetric (Fig. 1c). The magnetic field in the dynamo region is strong ($\Lambda \approx 0.8$) and dominated by small length scales without pronounced axisymmetry (Fig. 1f). The field at the planet's surface is large-scale, fairly axisymmetric, and has a mean strength of about 90 nT (Fig. 1g, h). In the dynamo region the spectrum of magnetic energy versus harmonic degree n has a maximum at $n = 7-10$ (Fig. 2a); the dipole ($n = 1$) is three times weaker. At the planet's surface the dipole and quadrupole are strongest and nearly equal, with the dipole dominating at some times (Fig. 1g) and the quadrupole at other times (Fig. 1h). However, both are weaker than would be expected from a geometric field decay with radius according to $r^{-(n+2)}$ from the top of the dynamo region to the planetary surface. For $n \geq 3$ the power in the surface magnetic field decreases rapidly with n (Fig. 2a).

More important than the geometric attenuation is the skin effect for controlling the strength and scale of the field outside the core. Let us ignore for simplicity the flow in the stable layer and assume that the dynamo-generated field diffuses through this region. It is time-dependent and, neglecting spherical geometry, is attenuated by the skin effect proportional to $\exp(-\Delta/[2\lambda\tau]^{1/2})$, where τ is the characteristic timescale of the field change and Δ the stable layer thickness. In the geomagnetic field τ varies with harmonic degree¹⁷ proportional to $n^{-1.35}$ and numerical dynamo models¹⁸ suggest that τ is controlled by the magnetic Reynolds number ($\tau \propto n^{-1} Rm^{-1}$). Because the model value of Rm is in the correct range, the characteristic times $\tau(n)$ should match those of Mercury's dynamo. Smaller

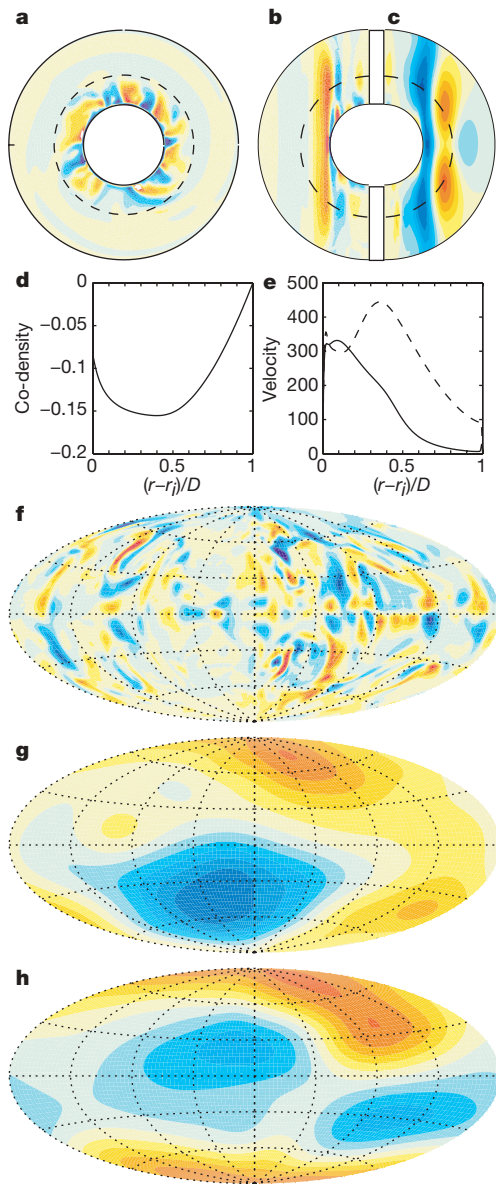


Figure 1 | Case I. The Prandtl number is $Pr = \nu/\kappa = 1$ and the magnetic Prandtl number is $Pm = \nu/\lambda = 3$. ν is viscosity and κ represents thermal and compositional diffusivity, the effective (turbulent) values of which are assumed to be equal in the co-density formulation. The inner core has the same electrical conductivity as the fluid shell and is co-rotating with the mantle. Boundary conditions are no-slip; the non-dimensional source term for co-density is $\varepsilon = -1.235$; the ICB flux is $\partial C/\partial r = -1$. The ratio of total core-mantle boundary co-density flux to ICB flux is -4 , as in the example in the text. Model results are scaled to physical values using $\lambda = 1.0 \text{ m}^2 \text{ s}^{-1}$ (other parameters given in the text). **a**, Snapshot of the vorticity component parallel to the rotation axis in the equatorial plane. **b**, Snapshot of vorticity in a cut along the rotation axis. **c**, Snapshot of mean flow in the azimuthal direction, with contour steps of $20\lambda/D$. Broken circles indicate the radius of neutral stability. **d**, Mean co-density profile versus radius. **e**, Poloidal (full line) and toroidal (broken line) velocity, normalized by λ/D and averaged in the horizontal direction and over time, versus radius. **f-h**, Radial magnetic field component, in **f** at $0.38D$ above the ICB with a 25,000 nT contour interval, and in **g** and **h** at the planetary surface ($r = 1.32r_o$) with 30 nT contours. Snapshots taken at $t = 108,000$ yr in **f** and **g** and 125,200 yr in **h**.

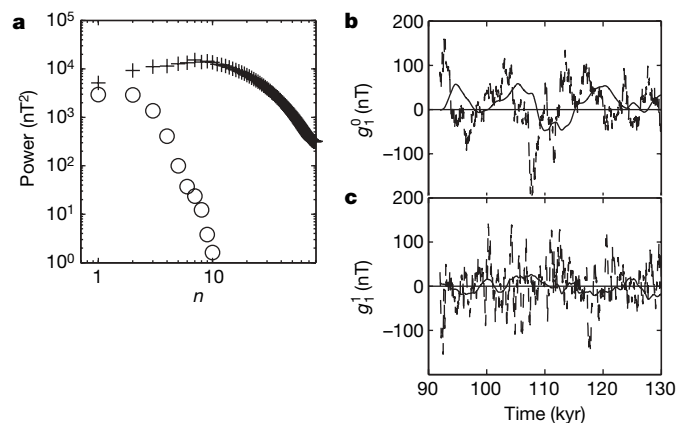


Figure 2 | Magnetic field spectral components of case I. **a**, Time-averaged magnetic power spectrum versus degree n at the planetary surface ($r = 1.32r_o$, circles) and radial average inside the fluid core (crosses). The latter values have been scaled down by a factor of 10^{-4} and represent mainly the field in the dynamo region. **b**, **c**, Time series of Gauss coefficients representing the axial dipole (**b**), and part of the equatorial dipole (**c**) at Mercury's surface (full lines). Broken lines are harmonic coefficients representing the poloidal magnetic field near the bottom of the stable layer at $r = r_i + D/2$, scaled down by a factor of 10^{-2} .

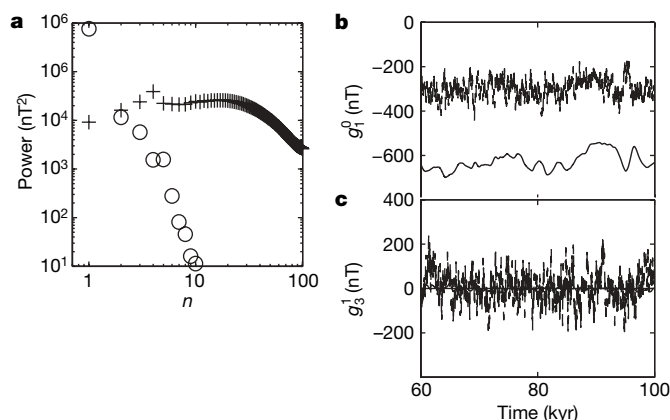


Figure 3 | Magnetic field spectral components of case II. $P_m = 3$, $P_r = 1$, and $\varepsilon = -1.5$. The unstable layer thickness is $0.4D$. **a, b**, As in Fig. 2. **c**, Time series of Gauss coefficient g_3^1 (as an arbitrary example for a multipole component) and equivalent coefficient for the field at $0.5D$ above the ICB (broken line).

scales in the magnetic field (higher n), which vary more rapidly with time, are more strongly damped by the skin effect.

Figure 2b shows the axial dipole field as function of time near the top of the dynamo region and outside the core. Short-term fluctuations of the dynamo field are eliminated, whereas slow variations with a characteristic period of $T = 10,000$ yr can pass to some degree. A phase shift of about $2,500$ yr is obvious. The skin effect predicts a shift of $T\lambda/(4\pi\lambda T)^{1/2} = 3,000$ yr for $\lambda = 600$ km and $\lambda = 1$ m² s⁻¹. The non-axial dipole components have amplitude similar to that of the axial part in the dynamo region (Fig. 2c), but little power at low frequency. Therefore, they are more strongly suppressed in the outside field. The differential rotation in the stable layer (Fig. 1c) also helps to attenuate non-axisymmetric components¹⁹. The axisymmetric modes contribute only 6% to the poloidal magnetic energy in the dynamo region, but carry—on time-average—two-thirds of the power of the field outside the core.

Because of the filtering effect the timescales of secular variation are larger for the outside magnetic field than for the dynamo field, $\tau(n) = 1,500, 1,000$ and 800 yr for $n = 1-3$, respectively. For comparison, the timescales of the Earth's quadrupole and octupole field are 160 and 240 yr, respectively¹⁷, although that for the dipole is larger.

In case I the mean surface magnetic field strength is lower by a factor of five than that given by the Mariner 10 observations. In case II, with a larger inner core ($r_i/r_o = 0.5$), the surface field has twice the observed strength. The power spectrum of the internal magnetic field is similar to that of case I, but the external field is now strongly dominated by the axial dipole (Fig. 3a). Compared to case I the dipole is more stable in time and did not reverse during the 100,000 yr of the model run (Fig. 3b). Therefore, it is not damped by the skin effect, in contrast to the multipole components that fluctuate on short timescales (Fig. 3c). The enhanced dipole stability in case II is possibly due to the larger magnetic inertia of the conducting inner core^{10,20}, whose dipole field can only change on timescales of more than $r_i^2/(\pi^2\lambda) = 2,700$ yr. This is significantly longer than the characteristic time of the intrinsic dynamo field fluctuations (Fig. 3b); therefore, the dipole magnetic flux anchored in the inner core (and in the stable layer of the fluid core) may stabilize the polarity in the dynamo region. The two cases bracket the observed field strength and their field geometry is compatible with our limited knowledge of Mercury's magnetic field. Conceivably the amplitude could be matched by adapting the stable layer thickness, the inner core radius or the buoyancy flux.

The Messenger and Bepi Colombo missions to Mercury will characterize its magnetic field much better. The following model predictions can be tested with new data. (1) Mercury's dipole does not necessarily dominate over the quadrupole, but the core field carries little energy at $n > 3$. A potential problem is that the core field may be obscured by crustal magnetization at higher n . (2) The magnetic field is predominantly axisymmetric. (3) No significant secular variation occurs on a decadal timescale. (4) The solid inner core must not be too large; otherwise the stable layer thickness would be insufficient for an efficient skin effect. A large inner core can potentially be detected by geodetic observations of Mercury's tides²¹ or libration²². A radius significantly larger than 1,000 km would falsify the deep dynamo model.

Received 27 August; accepted 9 October 2006.

1. Stevenson, D. J. Planetary magnetic fields. *Earth Planet. Sci. Lett.* **208**, 1–11 (2003).
2. Schubert, G., Ross, M. N., Stevenson, D. J. & Spohn, T. Mercury's thermal history and the generation of its magnetic field. In *Mercury* (eds Vilas, F., Chapman, C. R. & Matthews, M. S.) 429–460 (Univ. Arizona Press, Tucson, 1988).
3. Hauck, S. A., Dombard, A. J., Phillips, R. J. & Solomon, S. C. Internal and tectonic evolution of Mercury. *Earth Planet. Sci. Lett.* **222**, 713–728 (2004).
4. Ness, N. F. The magnetic field of Mercury. *Phys. Earth Planet. Inter.* **20**, 209–217 (1979).
5. Connerney, J. E. P. & Ness, N. F. Mercury's magnetic field and interior. In *Mercury* (eds Vilas, F., Chapman, C. R. & Matthews, M. S.) 494–513 (Univ. Arizona Press, Tucson, 1988).
6. Aharonson, O., Zuber, M. T. & Solomon, S. C. Crustal remanence in an internally magnetized non-uniform shell: a possible source for Mercury's magnetic field? *Earth Planet. Sci. Lett.* **218**, 261–268 (2004).
7. Christensen, U. R. & Aubert, J. Scaling properties of convection-driven dynamos in rotating spherical shells and application to planetary magnetic fields. *Geophys. J. Int.* **166**, 97–114 (2006).
8. Heimpel, M. H., Aurnou, J. M., Al-Shamali, F. M. & Gomez Perez, N. A numerical study of dynamo action as function of spherical shell geometry. *Earth Planet. Sci. Lett.* **236**, 542–557 (2005).
9. Stanley, S., Bloxham, J., Hutchinson, W. E. & Zuber, M. T. Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth Planet. Sci. Lett.* **234**, 27–38 (2005).
10. Takahashi, F. & Matsushima, M. Dipolar and non-dipolar dynamos in thin shell geometry with implications for the magnetic field of Mercury. *Geophys. Res. Lett.* **33**, L10202, doi:10.1029/2006GL02579 (2006).
11. Turner, J. S. *Buoyancy Effects in Fluids* Ch. 8 (Cambridge Univ. Press, Cambridge, 1973).
12. Braginsky, S. I. & Roberts, P. H. Equations governing convection in Earth's core and the geodynamo. *Geophys. Astrophys. Fluid Dyn.* **79**, 1–97 (1995).
13. Lister, J. R. & Buffett, B. A. The strength and efficiency of thermal and compositional convection in the geodynamo. *Phys. Earth Planet. Inter.* **91**, 17–30 (1995).
14. Kutzner, C. & Christensen, U. R. Simulated geomagnetic reversals and preferred virtual geomagnetic pole paths. *Geophys. J. Int.* **157**, 1105–1118 (2004).
15. Olson, P. & Christensen, U. R. Dipole moment scaling for convection-driven planetary dynamos. *Earth Planet. Sci. Lett.* **250**, 561–571 (2006).
16. Zhang, K. & Schubert, G. Teleconvection: remotely driven thermal convection in rotating stratified spherical layers. *Science* **290**, 1944–1947 (2000).
17. Olsen, N. *et al.* CHAOS—a model of the Earth's magnetic field derived from CHAMP, Ørsted, and SAC-C magnetic satellite data. *Geophys. J. Int.* **166**, 67–75 (2006).
18. Christensen, U. R. & Tilgner, A. Power requirement of the geodynamo from ohmic losses in numerical and laboratory dynamos. *Nature* **429**, 169–171 (2004).
19. Stevenson, D. J. Reducing the non-axisymmetry of a planetary dynamo and an application to Saturn. *Geophys. Astrophys. Fluid Dyn.* **21**, 113–127 (1982).
20. Hollerbach, R. & Jones, C. A. Influence of the Earth's inner core on geomagnetic fluctuations and reversals. *Nature* **365**, 541–543 (1993).
21. Van Hoolst, T. & Jacobs, C. Mercury's tides and internal structure. *J. Geophys. Res.* **108**, 5121, doi:10.1029/2003JE002126 (2003).
22. Peale, S. J., Phillips, R. J., Solomon, S. C., Smith, D. E. & Zuber, M. T. A procedure for determining the nature of Mercury's core. *Meteorit. Planet. Sci.* **37**, 1269–1283 (2002).

Acknowledgements Constructive reviews by S. Stanley and D. Stevenson are appreciated.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to U.R.C. (christensen@mps.mpg.de).

Observation of the radiative decay mode of the free neutron

Jeffrey S. Nico¹, Maynard S. Dewey¹, Thomas R. Gentile¹, H. Pieter Mumm¹, Alan K. Thompson¹, Brian M. Fisher², Isaac Kremsky², Fred E. Wietfeldt², Timothy E. Chupp³, Robert L. Cooper³, Elizabeth J. Beise⁴, Kristin G. Kiriluk⁴, James Byrne⁵ & Kevin J. Coakley⁶

The theory of quantum electrodynamics (QED) predicts that beta decay of the neutron into a proton, electron and antineutrino should be accompanied by a continuous spectrum of soft photons. While this inner bremsstrahlung branch has been previously measured in nuclear beta and electron capture decay, it has never been observed in free neutron decay. Recently, the photon energy spectrum and branching ratio for neutron radiative decay have been calculated using two approaches: a standard QED framework^{1–3} and heavy baryon chiral perturbation theory⁴ (an effective theory of hadrons based on the symmetries of quantum chromodynamics). The QED calculation treats the nucleons as point-like, whereas the latter approach includes the effect of nucleon structure in a systematic way. Here we observe the radiative decay mode of free neutrons, measuring photons in coincidence with both the emitted electron and proton. We determined a branching ratio of $(3.13 \pm 0.34) \times 10^{-3}$ (68 per cent level of confidence) in the energy region between 15 and 340 keV, where the uncertainty is dominated by systematic effects. The value is consistent with the predictions of both theoretical approaches; the characteristic energy spectrum of the radiated photons, which differs from the uncorrelated background spectrum, is also consistent with the calculated spectrum. This result may provide opportunities for more detailed investigations of the weak interaction processes involved in neutron beta decay.

The neutron is composed of two down quarks and an up quark and is stable under the strong and electromagnetic interactions. The weak interaction, however, can convert a down quark into an up quark through the emission of a virtual W gauge boson that subsequently decays into an electron and an antineutrino. In QED, the decay is also accompanied by an inner-bremsstrahlung photon in the process:

$$n \rightarrow p + e^- + \bar{\nu}_e + \gamma.$$

QED takes into account the inner bremsstrahlung produced by the electron while the heavy baryon chiral perturbation theory approach includes the photon emission from the weak interaction vertex. Because direct emission from the weak vertex contributes less than 1% to the total intensity, both the photon energy spectrum and the polarization are dominated by the electron inner bremsstrahlung. The total intensity of inner bremsstrahlung diverges logarithmically as the photon energy E goes to zero because the spectrum displays the $1/E$ behaviour that is characteristic of all soft photon processes. However, it has long been established that this infrared divergence is cancelled in all orders of perturbation theory by higher-order virtual photon corrections to the radiationless mode^{5–7}.

The experimental challenge for the detection of radiative neutron decay is to distinguish the low rate of radiative decay events at observable energies from the intense photon background associated with a neutron beam. The estimated branching ratio above 15 keV is only about 3×10^{-3} , which, when coupled with the long neutron lifetime, makes the rate of detectable photons quite small. A previous experimental study of neutron radiative decay arrived at an upper limit of 6×10^{-3} at the 90% confidence level for the branching ratio for emission of photons with an energy between 35 and 100 keV (ref. 8). The report of a more recent experiment⁹ was disputed¹⁰ with compelling arguments. Our experiment was mounted at the NG-6 fundamental physics end-station at the Center for Neutron Research at the National Institute of Standards and Technology. It was designed to detect the coincidence of a photon and electron followed by a delayed proton, thereby reducing the probability of recording uncorrelated background events¹¹. A strong magnetic field transported the electrons and protons away from the photon detector, which increased the solid angle for detection and minimized correlated backgrounds. The detection method used a solenoid design and proton detection scheme previously used to measure both the neutron lifetime^{12,13} and the electron–antineutrino angular correlation coefficient¹⁴. An electrostatic mirror was used to vary the rate of detected electron–proton coincidences without changing the uncorrelated photon background rate, thus providing a signature for the detection of radiative decay and an important systematic check on possible backgrounds.

The NG-6 cold neutron beam entered beryllium-coated guides and collimation originally designed for use in a neutron time-reversal violation experiment¹⁵ and adapted for this experiment by implementing slight changes in the beam optics. The collimation used a series of ⁶LiF apertures backed with lead to define the beam, and the vacuum components were lined with ⁶Li-glass to absorb scattered neutrons, thus significantly reducing background radiation. The beam entered parallel to the 4.6 T magnetic field produced by a superconducting solenoid. The detection scheme for the decay products is shown in Fig. 1. To satisfy the need for a large solid-angle photon detector that can operate in a strong magnetic field and at low temperatures, a system consisting of a scintillating crystal coupled to an avalanche photodiode was employed¹⁶. The photon was detected by a single bismuth germanate (BGO) crystal viewed by a silicon avalanche photodiode (APD). As the temperature decreases, the APD gain increases, its noise decreases¹⁷, and the BGO light output increases¹⁸; these features allowed us to obtain a low-energy detection threshold of 15 keV. The BGO crystal was mounted in an aluminium

¹Physics Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA. ²Department of Physics, Tulane University, New Orleans, Louisiana 70118, USA. ³FOCUS and Physics Department, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁴Department of Physics, University of Maryland, College Park, Maryland 20742, USA. ⁵Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK. ⁶Information Technology Laboratory, National Institute of Standards and Technology, Boulder, Colorado 80305, USA.

holder and placed below the neutron beam in the downstream end of the bore of the solenoid. The placement of the photon detector well away from the surface barrier detector (SBD) significantly reduced correlated background of external bremsstrahlung from electrons striking the SBD.

The preamplifier signal from the SBD was transmitted from high voltage via an analogue fibre optic link where it was split and amplified. Two single-channel-analyser windows were set to encompass the entire proton pulse-height peak and the electron peak just above the maximum proton peak height, so that the electron signal could not be triggered by a proton. The fast timing outputs provided the start (electron window signal) and stop (proton window signal) of a time-to-amplitude converter. If a proton stopped the time-to-amplitude converter within a 20 μ s timing window, a conversion signal triggered a computer-based digital oscilloscope board that recorded the amplified electron–proton coincidence signal and the preamplifier output of the APD. Figure 2 shows a histogram of electron–photon delay times and an example of the SBD and APD waveforms.

Histograms were made of both the energy and timing spectra, and a minimal number of software cuts were placed on the data. For the proton, an energy window of twice the full width at half maximum of the gaussian energy peak was used. The entire electron spectrum above the hardware threshold of approximately 35 keV was accepted. The electron–proton time window was 2.5 to 20 μ s, where the lower limit was chosen to eliminate the influence of tails from start pulses due to events other than electrons. Studies with the beam off confirm that such events, which make up approximately 10% of the electron–proton trigger rate before cuts, were typically due to beam-related γ -rays. In addition, a smaller number of events was rejected under the requirement that the start pulse should return to the baseline before accepting a stop event. Although the electron and radiative decay photon are nearly simultaneous, a window of approximately 20 μ s was placed on the time difference between the electron and photon to measure the uncorrelated background.

There are three main sources of correlated background that could produce a coincidence peak in the electron–photon timing spectrum: (1) external bremsstrahlung from a beta electron stopping in the SBD, (2) cascades of γ -rays that follow neutron capture on material near the detector, and (3) high voltage discharges. Bremsstrahlung originating in the SBD is nearly indistinguishable from a radiative decay event, but the geometry of the apparatus greatly reduces the

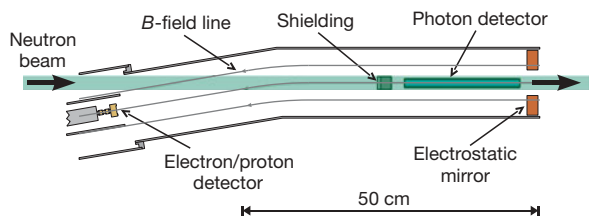


Figure 1 | Detection scheme for measuring the radiative decay of the neutron. The cold neutron beam traversed the bore of a superconducting solenoid. When a neutron decayed inside the high-field region, the charged decay products were confined to move in tight cyclotron orbits less than 1 mm in diameter, whose guiding centres followed the local magnetic field lines. The solenoid has a slight (9.5°) bend in the magnetic field direction at one end, allowing both the decay proton and electron to be guided out of the beam and into a silicon SBD 600 mm² in area and 1 mm thick. The SBD was held at a high negative potential (−25 kV) to accelerate the low-energy protons to detectable energy. A typical beta electron has a much higher velocity and reached the detector first, whereas the much slower proton (maximum energy of 751 eV) drifted to the detector a few microseconds later. The electrostatic mirror, constructed from an annulus through which the neutron beam passed, was used to reverse the direction of a proton with initial momentum directed away from the detector. Thus, the magnetic field and electrostatic mirror allowed for nearly 2π solid-angle coverage for electron detection and up to 4π coverage for proton detection. The photon detector and shielding lie below the neutron beam in the illustration.

probability of such photons reaching the BGO crystal. Because charged particles from neutron decay are constrained to tight orbits around the magnetic field lines, they cannot strike the bore or other material in the vicinity of the photon detector. They can strike only the SBD or the downstream end of the vacuum chamber, both of which are well separated from the photon detector. Furthermore, shielding composed of ^6Li -glass and 2 cm of lead partially occluded the direct line of sight between the two detectors. A simulation of the detector geometry was performed using the Monte Carlo code MCNP5 (ref. 19), and the possible contribution of external bremsstrahlung photons to the signal was determined to be $(3 \pm 3)\%$. Cascades of γ -rays that occur when a cold neutron is captured on material near the detectors can trigger both detectors. The characteristic waveforms of these events were determined by acquiring data with the accelerating voltage off, and this information was used to eliminate this source of background from the data analysis. Another possible source was high voltage discharge at the SBD. These waveforms were characterized with the beam off, and this information was also used to eliminate any significant contribution to the signal.

The experiment began in January 2004, and 23 runs with varying mirror potentials were performed from June to the end of November 2005. Some of the typical rates are given in Table 1. The radiative decay detection rate does not change linearly with the mirror potential, in contrast with non-bremsstrahlung sources of correlated backgrounds. A Monte Carlo simulation was written to calculate the expected ratio $R_{\text{ep}\gamma}/R_{\text{ep}}$ as a function of mirror potential. It includes the detector geometry, four-body decay phase space of the neutron,

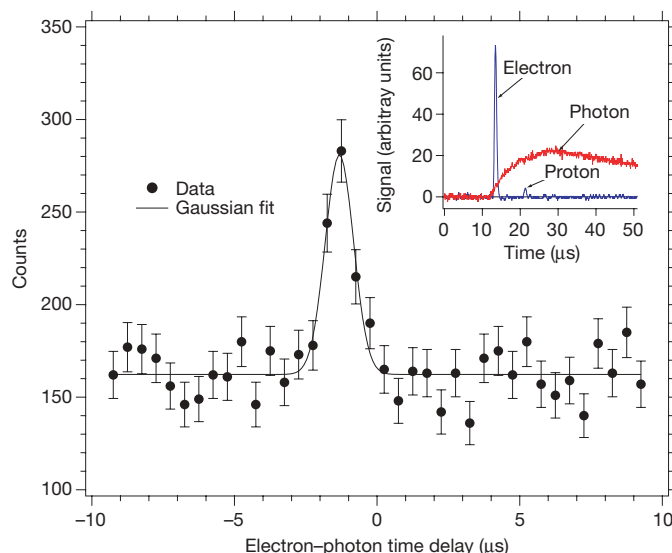


Figure 2 | Electron–photon timing spectrum for a three-day run with the mirror reflecting all protons. The spectrum shows all photons in a 20 μ s window for which the electron start pulse was accompanied by a delayed proton event. The error bars are statistical only and the confidence interval corresponds to 68%. In the inset, the blue line shows the SBD signal for an electron event followed by a delayed proton; the red line shows the much slower pre-amplified output of the APD for a photon event occurring in coincidence with the electron. Events attributable to radiative decay photons occur as prompt events in the spectrum and appear at -1.25μ s due to electronic delays. The approximately 1 μ s width of the peak arises from the uncertainty in extracting the onset time from the comparatively slow rise time of the APD signal. The flat background rate is consistent with a calculation of the random rate of coincidences based on the rates without any coincidence requirement and the time resolution. From the electron–proton waveform, the energy of each charged particle and the difference in arrival times between the electron and proton were extracted. The particle energies were obtained by integrating the pulses, and the timing was determined from the separation of the two peaks. The same parameters were obtained from the photon waveform by fitting the preamplifier signal to an empirically determined photon pulse shape.

Table 1 | Count rates determined from runs carried out at different mirror potentials

V_{mirror} (V)	Live time (d)	R_{ep} (s^{-1})	R_{γ} (s^{-1})	$R_{\text{ep}\gamma}$ (s^{-1})
0	36.1	5.1	0.020	$(15.5 \pm 1.7) \times 10^{-5}$
100	8.8	8.1	0.030	$(29.2 \pm 4.5) \times 10^{-5}$
200	8.3	10.8	0.045	$(39.3 \pm 5.1) \times 10^{-5}$
300	5.8	15.0	0.059	$(77.8 \pm 7.6) \times 10^{-5}$
400	5.6	17.4	0.068	$(76.4 \pm 8.0) \times 10^{-5}$
500	9.8	18.0	0.062	$(83.2 \pm 6.3) \times 10^{-5}$
700	2.5	20.1	0.072	$(124 \pm 14) \times 10^{-5}$
>750	11.7	20.2	0.076	$(94.0 \pm 6.2) \times 10^{-5}$

R_{ep} is the average rate of valid electron–proton triggers; R_{γ} is the average rate of photon events that have a valid electron–proton trigger within a 51 μs window; and $R_{\text{ep}\gamma}$ is the average rate of electron–photon events with a valid proton after background subtraction (that is, the rate of events in the peak of Fig. 2). The confidence interval corresponds to 68%.

and particle transport. For radiative decay events, the neutron decay kinematics and the detector acceptance yield a dependence on the mirror potential, as shown in Fig. 3. The drop in the ratio as the mirror potential goes to zero is due to the kinematics of neutron decay with photon emission. When the mirror potential is zero (or very low), only electrons and protons whose momenta have a component directed towards the SBD will be detected. Thus, the momenta of the antineutrino and emitted photon will tend to be in the opposite direction, away from the photon detector. When the mirror potential is raised above the maximum kinetic energy of the proton, the SBD detects protons emitted in all directions, and hence the photon momentum direction is less constrained.

The Monte Carlo simulation was used to extract the branching ratio, which is a single parameter that sets the scale of the ratio $R_{\text{ep}\gamma}/R_{\text{ep}}$. The data in Fig. 3 were fitted to the Monte Carlo with only a single multiplicative factor, so that data taken at all mirror potentials contribute to the determination of the branching ratio. The error bars indicate statistical uncertainties only, that is, uncertainties derived from random processes. The result, with its uncertainty from the fit, is $(3.13 \pm 0.11) \times 10^{-3}$ in the energy region between 15 and 340 keV, and the χ^2 value per degree of freedom is equal to 13.8/7.

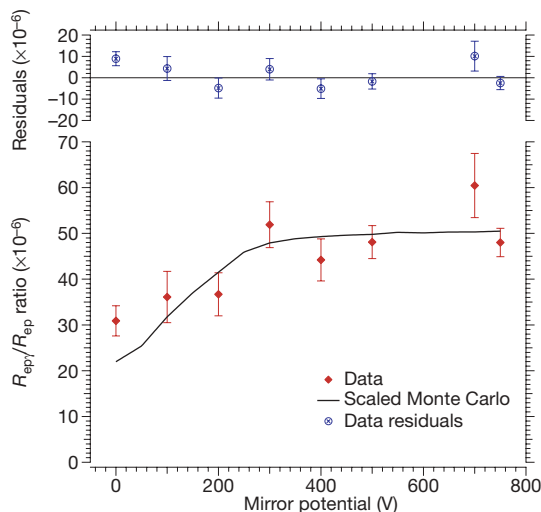


Figure 3 | Plot of the ratio $R_{\text{ep}\gamma}/R_{\text{ep}}$ versus the applied mirror potential. For each run, $R_{\text{ep}\gamma}$ was obtained from the number of events in the electron–photon timing peak. This value was divided by R_{ep} after subtracting the background due to beam-related triggers. The runs at particular mirror potentials were repeated several times. The final value was determined by a weighted average of the individual runs. The $R_{\text{ep}\gamma}/R_{\text{ep}}$ ratio for the runs performed at a given potential was reproducible within statistical uncertainty, even though most runs were separated by periods of many weeks. The solid line is the result of a Monte Carlo simulation that is scaled to fit the data points so as to determine the branching ratio. The error bars indicate statistical uncertainties only. The top trace shows the residuals of the fit.

The confidence intervals quoted here and throughout are 68%. The measured energy spectrum, shown in Fig. 4, is consistent with theoretical calculations^{1,4}. No correction has been applied to these data to incorporate the photon detector response function, but modelling indicates that this correction will be small in comparison with the uncertainties of this measurement.

The dominant systematic uncertainties are associated with the gain drift and energy calibration of the photon detector and the accuracy in determining the analysis windows, which contribute a relative standard uncertainty of 7.8%. Additional contributions result from the inputs of the Monte Carlo simulation, which include the spatial registration of both electric and magnetic fields relative to the detectors and the beam profile over the fiducial volume of the detector. These were varied in the Monte Carlo to yield an estimated uncertainty of 3.3%. In addition to the correction made for external bremsstrahlung, a correction was made for the photon detector efficiency and response of $(-3.0 \pm 3.0)\%$. Other systematics include electron backscattering from the SBD, magnitude of the magnetic field, and statistical uncertainty from the Monte Carlo. Combining all the systematic corrections and uncertainties in quadrature gives a total contribution of $(0 \pm 10.5)\%$ to the branching ratio.

This measurement represents the first observation of the photons associated with the radiative decay of the neutron. The branching ratio was measured to be $(3.13 \pm 0.34) \times 10^{-3}$ and is consistent with calculations from heavy baryon chiral perturbation theory (and QED) that predict a branching ratio of 2.85×10^{-3} in the same energy region (S. Gardner, personal communication). The total uncertainty is dominated by the contribution from systematic effects of 0.33×10^{-3} , while the statistical uncertainty contributes 0.11×10^{-3} . The photon energy spectrum is consistent with theoretical predictions, and the behaviour of the $R_{\text{ep}\gamma}/R_{\text{ep}}$ ratio as a function of the mirror potential coincides with the Monte Carlo prediction for radiative decay. Although the current result is limited by systematic effects, none of them presents a significant obstacle to

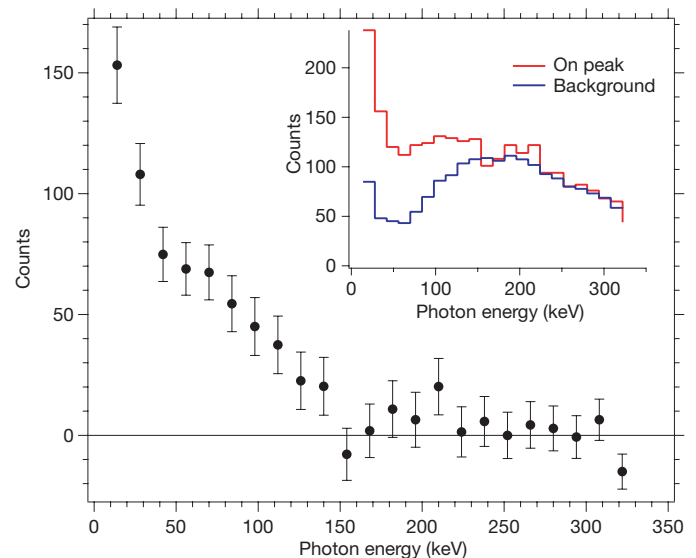


Figure 4 | Energy spectrum of photons from radiative neutron decay. The energy spectrum between 15 and 340 keV in the electron–photon timing peak was determined by subtracting the energy spectrum of the background from that of the peak, both shown in the inset. The peak spectrum was obtained by taking a 2 μs window centred around the peak of Fig. 2, and the background spectrum was determined from an 8 μs window in both the pre-prompt and post-prompt regions. The error bars indicate statistical uncertainties that were propagated from the difference of the two spectra. Calibrations were obtained using the 60 keV line from ^{241}Am and the 511 keV peak from electron–positron annihilation in the background spectrum. The spectrum shown here uses only runs in which the mirror voltage V_{mirror} was greater than 750 V (see Table 1).

improving the precision. Only a small fraction of the solid angle available for photon detection was used in this experiment. A 12-element scintillation detector is under construction that should permit a precision measurement of the photon spectrum and branching ratio at a level of a few per cent. A measurement of the spectrum below the 1% level could reveal direct emission from the weak vertex. Furthermore, a measurement of the photon circular polarization could reveal information about the Dirac structure of the weak current^{3,4}.

Received 31 August; accepted 25 October 2006.

- Gaponov, Yu. V. & Khafizov, R. U. Study of radiative neutron beta decay. *Phys. Atom. Nucl.* **59**, 1213–1220 (1996).
- Gaponov, Yu. V. & Khafizov, R. U. Radiative neutron β -decay and its possible experimental realization. *Phys. Lett. B* **379**, 7–12 (1996).
- Gaponov, Yu. V. & Khafizov, R. U. The radiative beta-decay mode of the free neutron. *Nucl. Instrum. Meth. A* **440**, 557–561 (2000).
- Bernard, V., Gardner, S., Meißner, U.-G. & Zhang, C. Radiative neutron β -decay in effective field theory. *Phys. Lett. B* **593**, 105–114 (2004); erratum **599**, 348 (2004).
- Jauch, J. M. & Rohrlich, F. *The Theory of Photons and Electrons* (Addison-Wesley, Cambridge, Massachusetts, 1955).
- Kinoshita, T. Mass singularities of Feynman amplitudes. *J. Math. Phys.* **3**, 650–677 (1962).
- Lee, T. D. & Nauenberg, M. Degenerate systems and mass singularities. *Phys. Rev. B* **133**, 1549–1562 (1964).
- Beck, M. *et al.* The upper limit of the branching ratio for radiative beta decay of free neutrons. *JETP Lett.* **76**, 332–336 (2002).
- Khafizov, R. U. *et al.* Observation of the neutron radiative decay. *JETP Lett.* **83**, 5–9 (2006).
- Severijns, N., Zimmer, O., Wirth, H.-F. & Rich, D. Comment on “Observation of the neutron radiative decay” by R. U. Khafizov *et al.* in *JETP Letters* **83**, 5–9 (2006). *JETP Lett.* **84**, 231 (2006).
- Fisher, B. M. *et al.* Detecting the radiative decay mode of the neutron. *J. Res. Natl. Inst. Stand. Technol.* **110**, 421–425 (2005).
- Nico, J. S. *et al.* Measurement of the neutron lifetime by counting trapped protons in a cold neutron beam. *Phys. Rev. C* **71**, 055502 (2005).
- Dewey, M. S. *et al.* Measurement of the neutron lifetime using a proton trap. *Phys. Rev. Lett.* **91**, 152302 (2003).
- Byrne, J. *et al.* Determination of the electron–antineutrino angular correlation coefficient a_0 and the parameter $|\lambda| = |G_A/G_V|$ in free neutron β -decay from measurements of the integrated energy spectrum of recoil protons stored in an ion trap. *J. Phys. G* **28**, 1325–1349 (2002).
- Mumm, H. P. *et al.* emiT: An apparatus to test time reversal invariance in polarized neutron decay. *Rev. Sci. Instrum.* **75**, 5343–5355 (2004).
- Gentile, T. R. *et al.* Particle and photon detection for a neutron radiative decay experiment. *Nucl. Instrum. Meth. A* (in the press).
- Yang, L. *et al.* Performance of a large-area avalanche photodiode at low temperature for scintillation detection. *Nucl. Instrum. Meth. A* **508**, 388–393 (2003).
- Piltingsrud, H. V. The low-temperature scintillation properties of bismuth germanate and its application to high-energy gamma radiation imaging devices. *J. Nucl. Med.* **20**, 1279–1285 (1979).
- Brown, F. B. *et al.* *A General Monte Carlo N-Particle Transport Code*. Version 5, LANL publication LA-UR-03-1987 (Los Alamos National Laboratory, 2003).

Acknowledgements We thank S. Gardner of the University of Kentucky for discussions and her interest in this work. We acknowledge the support of the National Institute of Standards and Technology, US Department of Commerce, which provided the neutron facilities used in this work. This research was made possible in part by support from the National Science Foundation and a US Department of Energy interagency agreement.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.S.N. (jnico@nist.gov).

Fortnightly variations in the flow velocity of Rutford Ice Stream, West Antarctica

G. Hilmar Gudmundsson¹

Most of the ice lost from the Antarctic ice sheet passes through a few fast-flowing and highly dynamic ice streams¹. Quantifying temporal variations in flow in these ice streams, and understanding their causes, is a prerequisite for estimating the potential contribution of the Antarctic ice sheet to global sea-level change^{2,3}. Here I show that surface velocities on a major West Antarctic Ice Stream, Rutford Ice Stream, vary periodically by about 20 per cent every two weeks as a result of tidal forcing. Tidally induced motion on ice streams has previously been thought to be limited to diurnal or even shorter-term variations^{4–9}. The existence of strong fortnightly variations in flow demonstrates the potential pitfalls of using repeated velocity measurements over intervals of days to infer long-term change.

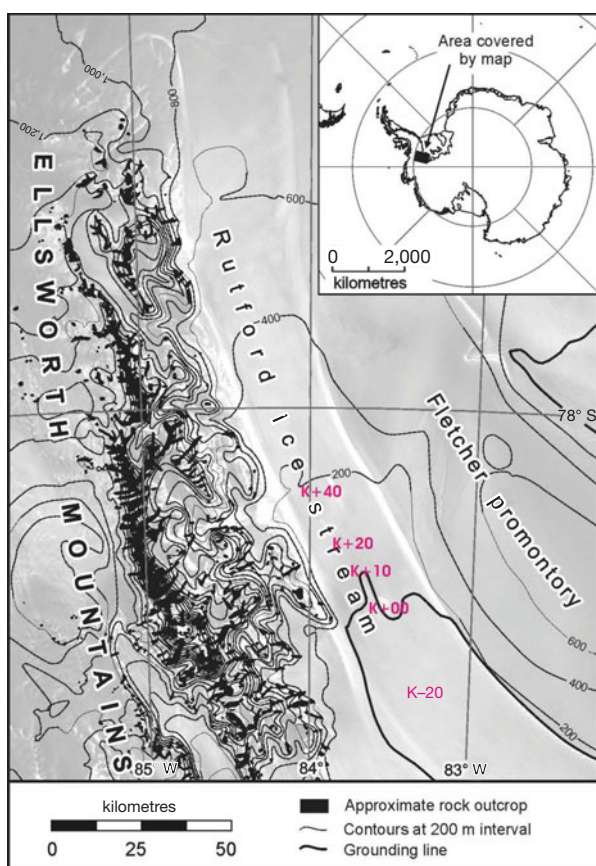


Figure 1 | Map of Rutford Ice Stream, West Antarctica. Locations of the GPS stations discussed in the text are indicated using the labels K – 20 to K + 40. The position of the grounding line is shown as a heavy black line.

Field measurements on Rutford Ice Stream (78° S, 83° W), West Antarctica (Fig. 1), have revealed fortnightly variations in flow speeds with about 20% peak-to-peak amplitude (Fig. 2). Global Positioning System (GPS) observations were collected continuously at five sites over a seven-week period from late December 2003 until mid-February 2004. One site was on a floating ice shelf some 20 km downstream of the grounding line. The other four locations were along the central flow line extending 40 km upstream of the grounding line (Fig. 1). Kinematic position solutions were calculated every 5 min using the Bernese GPS software^{10,11}.

The bottom panel of Fig. 2 shows hourly averaged velocities at four sites labelled R – 20, R + 00, R + 20, and R + 40, where numbers refer to the distance in kilometres upstream from the grounding line. (R – 20 was 20 km downstream of the grounding line.) For clarity, data from R + 10 are not shown in Fig. 2 (but are shown in Fig. 3). At all locations, fortnightly variations in flow speed are observed, but the peak-to-peak amplitudes of the velocity variations change with distance from the grounding line. At the grounding line, velocities change over one week from about 0.9 to 1.2 m day^{–1}, 27% of the total velocity. At R + 40 the range is 0.95 to 1.07 m day^{–1}, or about 12% variation. No significant temporal variations in vertical or transverse velocities were observed upstream of the grounding line.

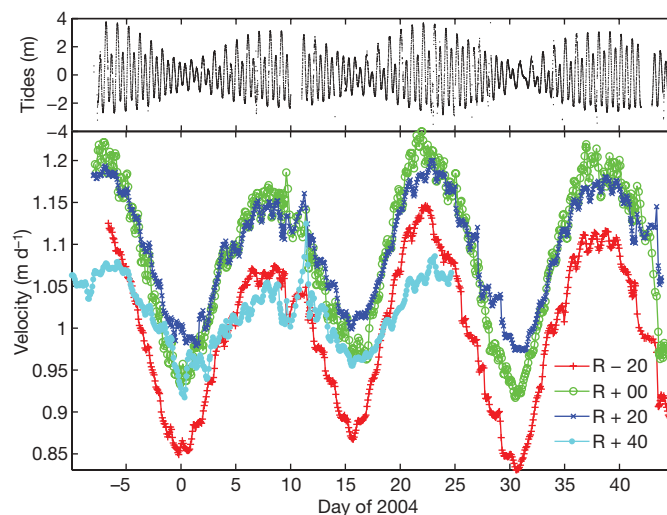


Figure 2 | Surface speed and tidal amplitudes as a function of time. The top panel shows vertical tidal amplitudes some 20 km downstream of the grounding line of the Rutford Ice Stream, as measured on the surface of the floating ice shelf. The bottom panel shows surface speeds 20 km downstream (red crosses), at the grounding line (green circles), and 20 km (dark blue exes) and 40 km (light blue circles) upstream from the grounding line. (The grounding line marks the division between the floating and the grounded parts of the ice.)

¹British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

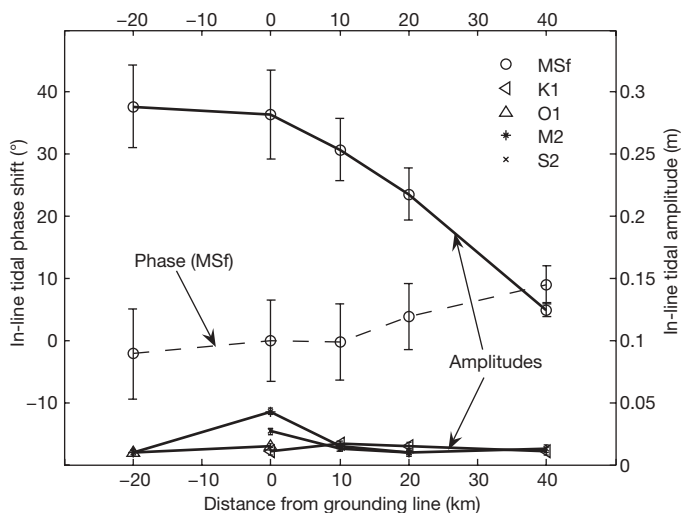


Figure 3 | Tidal components. Tidal amplitudes (thick solid lines) of horizontal in-line positions as functions of distance from grounding line, with distance measured positively in upstream direction. In addition, the phase shift (thin dashed lines) of the MSf tide with respect to the grounding line is shown. The MSf is a fortnightly tide with a period of 13.66 days. The O1 and K1 are diurnal tides with periods of 25.82 and 23.93 hours, respectively, and M2 and S2 are semidiurnal tides with periods of 12.42 and 12.00 hours. The error bars are 95% confidence intervals calculated using the bootstrap method.

The upper panel of Fig. 2 shows measured vertical displacement at the Ronne Ice Shelf at site R – 20 measured with the same techniques. The dominant tides are semi-diurnal with clear spring-to-neap tidal cycles. A comparison of the upper and the lower panels of Fig. 2 suggests a simple correlation between temporal variations in velocities and the spring-to-neap tidal cycle, with velocities generally increasing in the period leading up to a spring tide, and decreasing thereafter until the next neap tide. Superimposed on the fortnightly flow variations are rather smaller-amplitude diurnal and semi-diurnal variations in flow speeds.

Figure 3 shows results of a tidal analysis of detrended in-line positions, in-line positions being defined as horizontal positions measured along the mean flow direction. The tidal analysis was done using the *t_tide* software package¹². At all locations the lunisolar synodic fortnightly tidal constituent (MSf) with a period of 13.66 days has the largest amplitude. The MSf tidal amplitude is equally large at R – 20 and R + 00, but decays roughly linearly with distance upstream. The dashed thin line in Fig. 3 shows the phase shift of the fortnightly MSf tide with respect to the grounding line. The phase changes by about 9° from R + 00 to R + 40 which, when taking into account the estimates of the phase error, corresponds to a velocity of 1 to 2 m s^{–1}.

The neap-to-spring increase in velocity and the subsequent spring-to-neap velocity decrease (Fig. 2), suggests that the velocity perturbation is related to peak-to-peak changes in the semi-diurnal tide, implying a nonlinear system response. A possible explanation for the observed fortnightly flow variations goes as follows: elastic stresses

set up by the tides perturb the basal shear stress distribution of the ice stream. Because the relationship between basal motion and shear basal stress is nonlinear, the increase in basal motion by a positive perturbation in shear stress is larger than the corresponding decrease from an equally large but negative shear stress perturbation. Over one tidal cycle, this imbalance results in a net additional forward displacement. The magnitude of this imbalance varies with changes in peak-to-peak tidal amplitude from the spring tide to the neap tide, resulting in fortnightly variation in flow speeds.

Interferometric synthetic aperture radar (InSAR) data has been used extensively to calculate surface flow velocities on ice streams using one-day or three-day intervals^{13–15}, with longer intervals rarely used because of loss of coherence. As shown here, ice stream speeds can vary significantly in response to the spring-to-neap tidal cycle. Caution must therefore be exercised in interpreting InSAR velocities on the basis of image pairs separated by a few days.

Received 18 April; accepted 9 November 2006.

1. Bentley, C. R. Rapid sea-level rise soon from West Antarctic Ice Sheet collapse? *Science* **275**, 1077–1078 (1997).
2. Vaughan, D. G. How does the Antarctic Ice Sheet affect sea level rise? *Science* **308**, 1877–1878 (2005).
3. Alley, R. B., Clark, P. U., Huybrechts, P. & Joughin, I. Ice-sheet and sea-level changes. *Science* **310**, 456–460 (2005).
4. Harrison, W. D., Echelmeyer, K. A. & Engelhardt, H. Short-period observations of speed, strain and seismicity on Ice Stream B, Antarctica. *J. Glaciol.* **39**, 463–470 (1993).
5. Bindshadler, R. A., King, M. A., Alley, R. B., Anandakrishnan, S. & Padman, L. Tidally controlled stick-slip discharge of a West Antarctic ice stream. *Science* **301**, 1087–1089 (2003).
6. Anandakrishnan, S., Voigt, D. E., Alley, R. B. & King, M. A. Ice stream D flow speed is strongly modulated by the tide beneath the Ross Ice Shelf. *J. Geophys. Res.* **30**, doi:10.1029/2002GL016329 (2003).
7. Bindshadler, R. A., Vornberger, P. L., King, M. A. & Padman, L. Tidally driven stick-slip motion in the mouth of Whillans Ice Stream, Antarctica. *Ann. Glaciol.* **36**, 263–272 (2003).
8. Anandakrishnan, S. & Bentley, C. R. Micro-earthquakes beneath Ice Streams B and C, West Antarctica: observations and implications. *J. Glaciol.* **39**, 455–462 (1993).
9. Anandakrishnan, S. & Alley, R. B. Tidal forcing of basal seismicity of ice stream C, West Antarctica, observed far inland. *J. Geophys. Res.* **102**, 15183–15196 (1997).
10. Hugentobler, U., Dach, R., Fridez, P. & Meindl, M. *Bernese GPS Software version 5.0* (<http://www.bernese.unibe.ch/>) (The Astronomical Institute of the University of Bern, 2005).
11. King, M. A. Rigorous GPS data-processing strategies for glaciological applications. *J. Glaciol.* **50**, 601–607 (2004).
12. Pawlowicz, R., Beardsley, B. & Lentz, S. Harmonic analysis including error estimates in MATLAB using T TIDE. *Comput. Geosci.* **28**, 929–937 (2002).
13. Goldstein, R. M., Engelhardt, H., Kamb, H. & Frolich, R. M. Satellite radar interferometry for monitoring ice sheet motion: Application to an Antarctic ice stream. *Science* **262**, 1525–1530 (1993).
14. Frolich, R. M. & Doake, C. S. M. Synthetic aperture radar interferometry over Rutford Ice Stream and Carlson Inlet. *J. Glaciol.* **44**, 77–92 (1998).
15. Joughin, I., Tulaczyk, S. & Bindshadler, R. Changes in west Antarctic ice stream velocities: Observations and analysis. *J. Geophys. Res.* **107**, doi:10.1029/2001JB001029 (2002).

Acknowledgements I thank R. C. A. Hindmarsh for discussions and E. McGough for assistance with field work. This work was supported by a grant from the NERC GEF.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (ghg@bas.ac.uk).

Intrasexual competition and sexual selection in cooperative mammals

T. H. Clutton-Brock^{1,2}, S. J. Hodge¹, G. Spong^{1,3}, A. F. Russell^{1,4}, N. R. Jordan¹, N. C. Bennett², L. L. Sharpe⁵ & M. B. Manser⁶

In most animals, the sex that invests least in its offspring competes more intensely for access to the opposite sex and shows greater development of secondary sexual characters than the sex that invests most^{1,2}. However, in some mammals where females are the primary care-givers, females compete more frequently or intensely with each other than males^{3–5}. A possible explanation is that, in these species, the resources necessary for successful female reproduction are heavily concentrated and intrasexual competition for breeding opportunities is more intense among females than among males. Intrasexual competition between females is likely to be particularly intense in cooperative breeders where a single female monopolizes reproduction in each group⁶. Here, we use data from a twelve-year study of wild meerkats (*Suricata suricatta*), where females show high levels of reproductive skew, to show that females gain greater benefits from acquiring dominant status than males and traits that increase competitive ability exert a stronger influence on their breeding success. Females that acquire dominant status also develop a suite of morphological, physiological and behavioural characteristics that help them to control other group members. Our results show that sex differences in parental investment are not the only mechanism capable of generating sex differences in reproductive competition and emphasize the extent to which competition for breeding opportunities between females can affect the evolution of sex differences and the operation of sexual selection.

Meerkats (*Suricata suricatta*) breed cooperatively in groups of 3–50 in the arid savannahs of southern Africa. In each group, a single dominant female (who usually either inherits her position in her natal group or acquires it as a founding member of a new group) monopolizes reproduction, producing up to four litters of 3–7 pups per year for up to ten years, though subordinate females breed occasionally⁷. Over 90% of pups born to dominant females are fathered by the group's dominant male, who is usually an immigrant or founding member of the group⁸. As in other mammals, breeding females invest more heavily in producing and rearing young than males, paying the costs of gestation and lactation as well as making a larger contribution than breeding males to provisioning dependent young^{9,10}. Subordinate females, too, contribute more to provisioning and guarding dependent young than subordinate males¹¹.

Although reproductive skew is pronounced in both sexes, female meerkats gain larger reproductive benefits from acquiring dominant status than males. The average number of pups surviving to 12 months produced by 33 dominant females during their period of tenure was 17.0 (range = 0–65), whereas the average number of pups fathered by 53 dominant males was 8.9 (range = 0–42; Mann–Whitney U test, $W = 2,070.5$; $N = 33, 53$; $P = 0.037$; Fig. 1a). In both

sexes, individual differences in breeding success depend primarily on the number of months for which individuals hold the dominant breeding position in their group (Fig. 1b), but the average tenure of females that acquire dominant status and breed at least once (31.5 months) is longer than that of males (17.4 months) (Hazard Ratio $\pm$ 95% Confidence Interval (CI) = 1.85 ± 0.31 ; $P = 0.041$; Fig. 1c). Sex differences in reproductive tenure occur partly because dominant females are less commonly replaced by competitors before their death and partly because dominant males are more likely to leave the group following the death of their partner than dominant females: although 80% of 'widows' retain their status for at least six months after the death of their partner and commonly breed with a

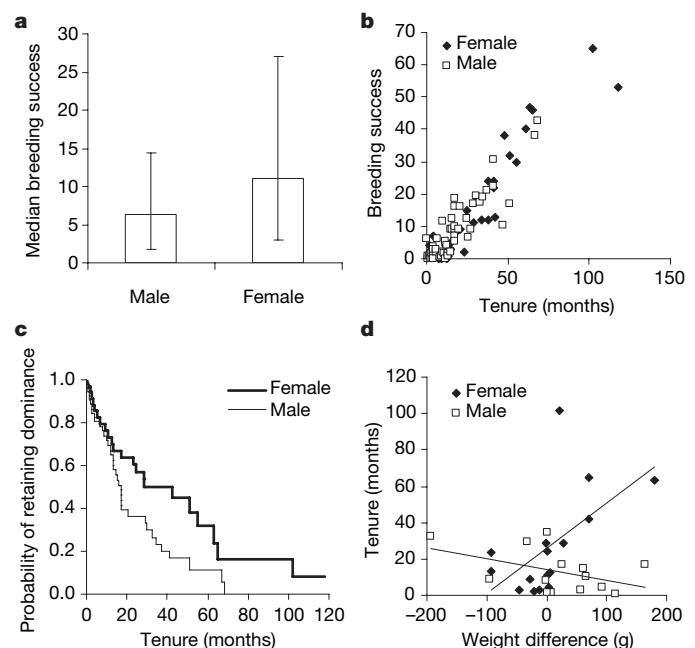


Figure 1 | Dominance and breeding success in males and females. **a**, Breeding success (offspring surviving to 12 months) of 33 dominant females and 53 dominant males. Graph shows medians and interquartile range. **b**, The influence of dominance tenure on breeding success in 33 dominant females (linear regression: $F_{1,32} = 226.44$, $P < 0.001$, $R^2 = 87.6$) and 53 dominant males (linear regression: $F_{1,52} = 185.91$, $P < 0.001$, $R^2 = 78.1$). **c**, Kaplan–Meier estimates of the probability of dominant females ($N = 33$) and males ($N = 53$) retaining their status after different periods of time. **d**, The relationship between the difference in body weight between the dominant and the heaviest same-sex competitor and tenure in dominant females ($N = 16$) and dominant males ($N = 14$).

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. ²Mammal Research Institute, University of Pretoria, 0002 Pretoria, South Africa.

³Department of Animal Ecology, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. ⁴Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK. ⁵Department of Zoology, University of Stellenbosch, Matieland 7602, South Africa. ⁶Zoologisches Institut, Universität Zürich, CH-8057 Zürich, Switzerland.

new partner, only 25% of 'widowers' do so (chi-square, $\chi^2_1 = 8.17$; $N = 15$ females, 12 males; $P < 0.004$).

As a result of the longer tenure of females, fewer females than males breed as dominants and standardized variance (variance/mean²) in estimates of lifetime breeding success is larger in females than males (females = 6.13; 95% CI = 3.66, 11.28; $N = 59$; males = 3.98; 95% CI = 2.23, 6.71; $N = 59$; see Supplementary Methods for details). Because the benefits of acquiring dominance status are greater for females than for males (Fig. 1a) females would be expected to compete more intensely with each other for dominant positions. As predicted, subordinate females of all ages initiate (Generalized Linear Mixed Model (GLMM): $\chi^2_1 = 10.09$; $P < 0.001$) and receive more aggression (GLMM: $\chi^2_1 = 19.77$; $P < 0.001$) than males of similar age and status (Fig. 2a, b). Subordinate females are also more frequently submissive than subordinate males (GLMM: $\chi^2_1 = 427.66$; $P < 0.001$; Fig. 2c). Though males and females do not differ in the frequency with which they win play fights as pups, as they grow older, females win a greater proportion of play fights than males (GLMM age*sex: $\chi^2_2 = 13.77$; $P = 0.001$; Fig. 2d).

Traits affecting competitive ability also have a stronger influence on the acquisition of dominance in females than in males. When subordinate females compete for the dominant position, older individuals usually win, but where there is no difference in age or younger individuals win, those that acquire dominant status are heavier than those that fail to do so (paired *t*-test: $t_{13} = 2.60$; $P = 0.023$; Fig. 3a). After the effects of body mass have been allowed for, daughters of dominant females are also more likely to acquire dominant status than those of subordinates (GLMM: $\chi^2_1 = 3.84$; $P = 0.050$; Fig. 3b). Older subordinate males usually win contests for dominant status but, where this is not the case, winners are not consistently heavier than losers (paired *t*-test: $t_{10} = 0.08$; $P = 0.94$; Fig. 3a) and the sons of dominant mothers are no more likely to acquire dominant status than those of subordinates (GLMM: $\chi^2_1 = 0.15$; $P = 0.70$; Fig. 3b), generating a significant interaction between the effects of sex and

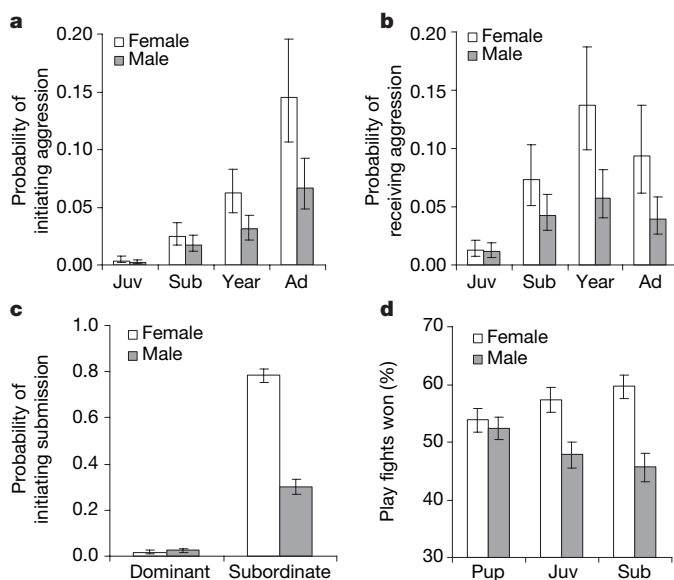


Figure 2 | Sex differences in aggression and submission. **a, b**, Probability that male and female subordinates of different age categories initiate or receive aggression during a breeding attempt (juvenile, 3–6 months; sub-adult, 6–12 months; yearling, 12–24 months; adult, >24 months). **c**, Probability that dominant and subordinate males and females >1 yr will initiate submission. **d**, The percentage of play fights involving opposite sexed pups (1–3 months), juveniles (3–6 months) and subadults (6–12 months) won by each sex ($N = 28$ females, 24 males); graphs show the predicted means ($\pm$ s.e.m) derived from GLMMs in which group, litter and individual identity were included as random terms. Sample sizes are provided in Methods.

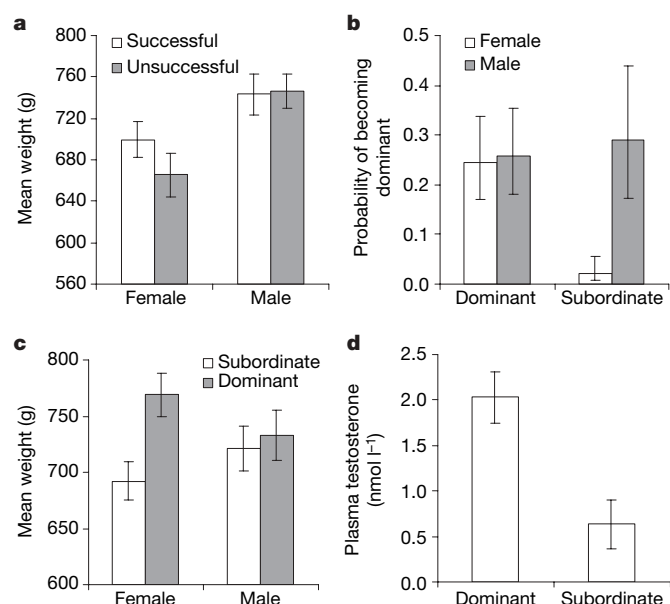


Figure 3 | The acquisition and maintenance of dominance status. **a**, Mean body mass in the three months before a dominance change of females ($N = 13$) and males ($N = 10$) that successfully acquired dominant status compared with an older, or same-aged unsuccessful competitor. **b**, Proportion of daughters and sons born to dominant ($N = 97$ daughters, 104 sons) or subordinate ($N = 23$ daughters, 20 sons) mothers that acquired dominant status. **c**, Mean mass (g) of females ($N = 13$) and males ($N = 8$) three months before becoming dominant, and six months after becoming dominant. **d**, Plasma testosterone concentrations (nmol l⁻¹) of dominant ($N = 12$) and subordinate ($N = 13$) adult females during pregnancy. All graphs show means $\pm$ s.e.m.

mother's dominance status on the probability of acquiring dominance status (GLMM: $\chi^2_1 = 6.62$, $P = 0.010$).

Females that acquire dominant status show a suite of behavioural and morphological changes that are reduced or absent in males. They increase in body mass whereas males do not (Fig. 3c), generating a significant interaction between the effects of dominance status on changes in body mass between the two sexes (Linear Mixed Model (LMM): $\chi^2_1 = 4.63$; $P < 0.031$). During pregnancy, dominant females also show higher levels of circulating testosterone than subordinates (General Linear Model (GLM): $F_{1,25} = 9.64$; $P = 0.005$; Fig. 3d), whereas males that acquire dominant status show no consistent increase in testosterone¹². Dominant females are also more frequently aggressive to other group members than dominant males (GLMM: $\chi^2_3 = 9.56$; $P = 0.021$), especially in late pregnancy when they direct aggression at older subordinate females, whom they commonly evict from their groups^{7,13}. These changes probably help to secure their position, for the tenure of dominant females increases with the difference between their average body weight and that of the heaviest subordinate female in the group (GLM: $F_{1,15} = 5.39$, $P = 0.037$). In contrast, there is no consistent association between the breeding tenure of dominant males and the difference between their weight and that of the heaviest subordinate male (GLM: $F_{1,13} = 1.89$, $P = 0.20$), generating a significant interaction between the effects of sex and relative body mass on dominance tenure in the two sexes (GLM: $F_{1,29} = 6.52$, $P = 0.017$; Fig. 1d).

Our results show that female meerkats that acquire dominant status have longer breeding tenures and gain greater reproductive benefits than males. A possible reason for the longer tenures of breeding females is that, once a dominant female is established, her potential competitors are natal animals that she can suppress or evict before they become serious rivals¹³, whereas the principal competitors of dominant males are either mature immigrants or members of other groups whose development they cannot control. Females are

also more likely to retain their position after the death of their partner because they commonly attract immigrant males from other groups, who are tolerated by their resident subordinate sons (who will eventually disperse to breed). In contrast, dominant males that lose their partners are not accepted as mates by natal females (who are commonly their daughters) and cannot attract unrelated female immigrants because these are not tolerated by resident subordinate females¹³. The greater benefits of acquiring dominance status may explain why females compete more intensely with each other for breeding opportunities throughout their lives and why traits enhancing competitive ability show greater development in dominant females than in dominant males.

Variation in female breeding success is unusually large in other cooperative or eusocial breeders⁶, and females are often more frequently aggressive than males. High levels of reproductive skew, long breeding lifespans and intense intrasexual competition for breeding opportunities are characteristic of females in many eusocial insects^{14,15}. Similarly, in naked mole-rats (*Heterocephalus glaber*), which live in colonies of up to a hundred or more with a single breeding female in each colony who breeds several times each year¹⁶, dominant females are more aggressive than other group members. Like dominant female meerkats, they show higher levels of testosterone than subordinates¹⁷ as well as a secondary period of growth after attaining dominant status, which is reduced or absent in males¹⁸. Although the reproductive tenure of dominant animals in natural populations of naked mole-rats is not known, captive breeding females can maintain their status for over twenty years¹⁹ and variance in female breeding success is likely to be high. Intense intrasexual competition combined with greater development of secondary sexual characters in females also occurs in a number of other vertebrates that do not breed cooperatively. In several of these species, successful reproduction depends on access to scarce or concentrated resources and females compete intensively with each other for these. For example, in spotted hyenas (*Crocuta crocuta*), where females are larger and more aggressive than males, females compete more intensely than males for social rank, which affects their access to resources as well as the survival of their dependent progeny³ but has little effect on breeding success in males²⁰. Dominant females develop high testosterone levels in late pregnancy and their offspring are more competitive²¹ and achieve higher status than those of subordinates²². Similarly, in clown fishes (*Amphiprion*), protandrous females compete for access to scarce sea anemones that provide safe breeding sites and are larger and more frequently aggressive than males^{23,24} and in *Eclectus* parrots (*Eclectus roratus*), females are brighter than males, and compete intensely for scarce nesting holes²⁵.

Our analyses suggest that, although greater parental investment by female mammals commonly leads to intense competition between males for access to mates and to strong selection favouring competitive ability in males^{1,26}, competition for reproductive opportunities between females can also generate strong selection pressures favouring competitive ability. In extreme cases, these may be capable of reversing the usual direction of sex differences in behaviour and morphology. Strong selection for competitive ability in females may sometimes be associated with greater variance in breeding success among females than among males⁶, but qualitative differences in the determinants of breeding success in males and females may also lead to divergent selective pressures on particular traits in the two sexes, and to sex differences in competitive behaviour and morphology²⁷. For example, the larger size and greater aggressiveness of females in spotted hyenas does not necessarily indicate that variation in breeding success is greater in females and could also occur because size and aggressiveness have a stronger effect on the breeding success of females than males. The relative benefits of successful competition are also likely to vary throughout the reproductive cycle in both sexes, and sex differences in competitiveness may also change²⁸. As a result, although sex differences in parental investment and associated indices (including the operational sex ratio, the opportunity for sexual

selection and comparative Bateman gradients) may usually predict the direction of sex differences in behaviour and morphology, they will not always do so and exceptions may be more common than is currently appreciated.

Finally, our results illustrate the limitations of current definitions of sexual selection. Although Darwin²⁹ sometimes described sexual selection as a consequence of intrasexual competition for breeding opportunities and sometimes as a consequence of intrasexual competition for mates, it is now commonly used to refer to selection operating through intrasexual competition for mates or mating opportunities^{2,30}. Unlike males, females more commonly compete with each other for opportunities to conceive or rear young (or for resources that allow them to do so, including male care) than for access to gametes produced by the opposite sex, so that current definitions of sexual selection¹ exclude the consequences of many important forms of intrasexual competition between females. An alternative approach is to return to Darwin's broader description of sexual selection as a process operating through intrasexual competition for reproductive opportunities, providing a conceptual framework that is capable of explaining the evolution of pronounced secondary sexual characters in females as well as in males.

METHODS

Study site. Data were collected at the Kuruman River Reserve, South Africa (26° 58' S, 21° 49' E), between October 1993 and May 2005. Descriptions of habitat, climate and study population are provided elsewhere¹³. All animals in the study population were habituated to close observation and most could be weighed before the morning foraging session without the need for capture.

Breeding success. The breeding success of 33 dominant females and 53 dominant males was measured as the number of pups surviving to 12 months produced during their tenure. As dominant males sire over 90% of pups born to dominant females and 45% of pups born to subordinate females (G.S. *et al.*, submitted) we assumed that 90% of pups born to dominant females and 45% of pups born to subordinate females were fathered by the resident dominant male during his tenure.

Statistical analysis. Multivariate analyses were performed in Genstat 5 (Release 4.2, Lawes Agricultural Trust, Rothamsted, Harpenden, UK). Where multivariate analyses were required, GLMs were used. Where necessary, LMMs and GLMMs were used to take into account repeated sampling within individuals, breeding attempts or groups. Random terms were retained in the model unless variance components were found to be zero. All statistical tests are two-tailed. Anderson-Darling tests were used to test for normality.

Dominance tenure. Tenure was measured from the date on which an individual acquired dominance to the date that dominance ended or observations ceased. Survival analysis (Cox's proportional hazards regression) was used to compare the probability of losing dominance for females ($N = 33$) and males ($N = 53$) from 21 social groups. To investigate the influence of relative weight at the start of the dominance period on the tenure of male ($N = 14$) and female ($N = 16$) dominants, dominance tenure was log transformed and fitted as the response term in a GLM. The differences in the mean non-breeding weights of dominants and their largest competitor in the two months around a dominance change, and the sex of the dominant individual, were fitted into the model as the main terms of interest. The number of competitors and group size at the start of dominance, and whether or not the dominant was the oldest individual in the group, were included as covariates but had no significant influence ($P > 0.05$).

Aggression and submission. We recorded all incidences of aggression and submission *ad libitum* during daily watches of groups during three-month periods following the birth of a litter. Aggression was defined as an occasion when an individual growled at, hip-slammed, bit or chased another individual. Submission was defined as an occasion when an individual emitted a peeping vocalisation, accompanied by a characteristic crouched posture. Whether or not a subordinate individual was seen to give or receive aggression from any other group member during a breeding attempt was fitted as the binomial response term (1 = Yes, 0 = No) in two separate GLMMs with 1 as the binomial denominator. The age category (juvenile, subadult, yearling or adult) and sex of the individual were fitted as the main terms of interest and group size was fitted as a covariate. In two further models, we fitted whether or not an individual older than 12 months gave aggression or submission to at least one other individual older than 12 months (of either sex) during a breeding attempt as the binomial response term (1 = Yes, 0 = No). The status of the individual (dominant male, dominant female, subordinate, male or subordinate female) was fitted as the main term of interest and group size was fitted as a covariate. Only individuals

who were present on all days on which data were collected were included in these analyses. Analyses of aggression were conducted on 1,457 periods of potential aggression, involving 335 individuals ($N = 42$ dominants and 324 subordinates) during 143 breeding attempts in 14 groups. Analyses of submission were conducted on 3,653 periods of potential submission involving 515 individuals ($N = 71$ dominants, 495 subordinates) in 374 breeding attempts and 23 groups. Group, individual and litter identity were included as random terms in all models.

Play fights. Bouts of play wrestling were recorded *ad libitum* between April 1999 and November 2002. For each wrestling bout, the individual that ended up on top was considered the 'winner'. We fitted the number of play fights won by each individual in each age category as the binomial response in a GLMM, with the total number of play fights observed as the binomial denominator. We included the age category of the individual (pup, juvenile or subadult) and the sex of the individual as fixed effects. Analysis was conducted on 52 individuals ($N = 28$ females, 24 males) from nine litters in seven groups, all of whom were observed in all three age categories. Individual, litter and group identity were included as random terms.

Influence of mother's dominance status. We included all offspring that survived to one year that could have reached four years of age before 01/04/05 in the analysis. Data were available from 97 daughters and 104 sons born to 15 dominant mothers, and 23 daughters and 20 sons and born to 19 subordinate mothers. Whether or not offspring attained dominance before 01/04/05 ($1 = \text{Yes}$, $0 = \text{No}$) was fitted as the binomial response term in a GLMM, with 1 as the binomial denominator. The sex of the offspring and the dominance status of the mother were included as the main fixed effects. Mean weight at 12 months (between 365 and 395 days) was included as a covariate, but was found to be non-significant ($P > 0.05$). Group and mother's identity were included as random terms.

Changes in body weight on attaining dominance. To investigate whether individuals became heavier after attaining dominance, we compared measures of body mass ($N = 103$) for 21 individuals ($N = 13$ females, 8 males) who acquired dominant status during the course of the study, using a LMM. Only the weights of non-pregnant females were included. Dominance status and sex were fitted as the main terms of interest. Age was included as a significant covariate (LMM: $\chi^2_1 = 16.18$, $P < 0.001$) and group size was included but was found to be non-significant ($P > 0.05$). Individual identity was included as a random term.

Hormonal analyses. Blood samples for hormonal analysis were collected from 25 females during the early stages of pregnancy ($N = 12$ dominant, 13 subordinate) living in 11 social groups between January 2006 and April 2006. Methods of hormonal analysis are described elsewhere¹². The testosterone concentration for each female (nmol l^{-1}) was fitted as the normally distributed response term in a GLM, with dominance status (dominant or subordinate) as the main term of interest. Oestrogen concentration (pmol l^{-1}) was included as a covariate and was found to have a significant positive influence (GLM: $F_{1,25} = 8.89$, $P = 0.007$). Female age and pregnancy stage (in days) were also included as covariates but were found to be non-significant ($P > 0.05$).

Received 26 July; accepted 23 October 2006.

1. Trivers, R. L. in *Sexual Selection and the Descent of Man, 1871–1971* (ed. Campbell, B.) 136–179 (Aldine-Atherton, Chicago, 1972).
2. Andersson, M. *Sexual Selection* (eds Krebs, J. R. & Clutton-Brock, T. H.) (University Press, Princeton, 1994).
3. Holekamp, K. E., Smale, L. & Szykman, M. Rank and reproduction in the female spotted hyaena. *J. Reprod. Fert.* **108**, 229–237 (1996).
4. Bennett, N. C. & Faulkes, C. G. *African Mole-rats: Ecology and Eusociality* (University Press, Cambridge, 2000).
5. Isaac, J. Potential causes and life-history consequences of sexual size dimorphism in mammals. *Mammal Rev.* **35**, 101–115 (2005).
6. Hauber, M. E. & Lacey, E. A. Bateman's principle in cooperatively breeding vertebrates: the effects of non-breeding alloparents on variability in female and male reproductive success. *Integr. Comp. Biol.* **45**, 903–914 (2005).
7. Clutton-Brock, T. H. *et al.* Cooperation, control and concession in meerkat groups. *Science* **291**, 478–481 (2001b).
8. Griffin, A. S., Pemberton, J. M., Brotherton, P. N. M., Gaynor, D. & Clutton-Brock, T. H. A genetic analysis of breeding success in the cooperative meerkat (*Suricata suricatta*). *Behav. Ecol.* **14**, 472–480 (2003).

9. Scantlebury, M., Russell, A. F., McIlrath, G. M., Speakman, J. R. & Clutton-Brock, T. H. The energetics of lactation in cooperatively breeding meerkats *Suricata suricatta*. *Proc. R. Soc. Lond. B* **269**, 2147–2153 (2002).
10. Clutton-Brock, T. H., Russell, A. F. & Sharpe, L. L. Behavioural tactics of breeders in cooperatively breeding meerkats. *Anim. Behav.* **68**, 1029–1040 (2004).
11. Clutton-Brock, T. H. *et al.* Evolution and development of sex differences in cooperative behavior in meerkats. *Science* **297**, 253–256 (2002).
12. Carlson, A. A. *et al.* Hormonal correlates of dominance in meerkats (*Suricata suricatta*). *Horm. Behav.* **46**, 141–150 (2004).
13. Clutton-Brock, T. H. *et al.* Infanticide and expulsion of females in a cooperative mammal. *Proc. R. Soc. Lond. B* **265**, 2291–2295 (1998b).
14. West-Eberhard, M. J. Sexual selection, social competition and speciation. *Q. Rev. Biol.* **55**, 155–183 (1983).
15. Shellman-Reeve, J. S. in *The Evolution of Social Behaviour in Insects and Arachnids* (eds Choe, J. C. & Crespi, B. J.) 52–93 (University Press, Cambridge, 1997).
16. Brett, R. A. in *Reproduction in Naked Mole-rats* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 67–96 (Princeton University Press, Princeton, 1991).
17. Faulkes, C. G. & Abbott, D. H. in *Cooperative Breeding in Mammals* (eds Solomon, N. G. & French, J. A.) 302–334 (University Press, Cambridge, 1997).
18. O'Riain, M. J., Jarvis, J. U. M., Alexander, R., Buffenstein, R. & Peeters, C. Morphological castes in a vertebrate. *Proc. Natl Acad. Sci. USA* **97**, 13194–13197 (2000b).
19. Sherman, P. W. & Jarvis, J. U. M. Extraordinary life spans of naked mole-rats (*Heterocephalus glaber*). *J. Zool.* **258**, 307–311 (2002).
20. Engh, A. L. *et al.* Reproductive skew among males in a female-dominated mammalian society. *Behav. Ecol.* **13**, 193–200 (2002).
21. Dloniak, S. M., French, J. A. & Holekamp, K. E. Rank-related maternal effects of androgens on behaviour in wild spotted hyenas. *Nature* **440**, 1190–1193 (2006).
22. Engh, A. & Holekamp, K. E. in *Social Complexity* (eds Waal, F. B. M. D. & Tyack, P.) (University Press, Harvard, 2002).
23. Fricke, H. W. & Fricke, S. Monogamy and sex change by aggressive dominance in coral fish. *Nature* **266**, 830–832 (1977).
24. Buston, P. Forcible eviction and prevention of recruitment in the clown anemone fish. *Behav. Ecol.* **14**, 576–582 (2003).
25. Heinsohn, R., Legge, S. & Endler, J. A. Extreme reversed sexual dichromatism in a bird without sex role reversal. *Science* **309**, 617–619 (2005).
26. Emlen, S. T. & Oring, L. W. Ecology, sexual selection, and the evolution of mating systems. *Science* **197**, 215–223 (1977).
27. Clutton-Brock, T. H. in *Evolution from Molecules to Men* (ed. Bendall, B. J.) 457–481 (University Press, Cambridge, 1983).
28. Forsgren, E., Amundsen, T., Borg, A. A. & Bjelvenmark, J. Unusually dynamic sex roles in a fish. *Nature* **429**, 551–554 (2004).
29. Darwin, C. *The descent of man and selection in relation to sex* (Murray, London, 1871).
30. Clutton-Brock, T. H. in *Sexual Selection in Primates: New and Comparative Perspectives* (eds Kappeler, P. M. & Schaik, C. P. V.) (University Press, Cambridge, 2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Kotze, J. Kotze and Northern Cape Conservation for allowing us to work in the Kalahari; G. McIlrath for his contribution to habituating meerkat groups; T. Flower, F. Dalerum, M. Haupt, J. d. Toit and E. Cameron of the University of Pretoria and P. Roth of the University of Cambridge for logistical support. We are grateful to the many volunteers, PhD students and post-docs who have assisted with data collection. We also thank A. Young, K. Isvaran and J. Hoogland for comments on previous drafts; P. Sherman and J. Shellman-Reeve for discussion; and the Natural Environmental Research Council, the Royal Society and the Earthwatch Institute for financial support.

Author Contributions T.H.C.B. and S.J.H. were responsible for the organisation and management of the project, drafting the manuscript and conducting most of the statistical analyses. Additional analyses were conducted by A.F.R. and L.L.S. G.S. was responsible for the genotyping and N.C.B. conducted all hormonal analyses. N.R.J. and M.B.M. contributed to the organisation and management of data collection in the field.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.H.C.B. (thcb@cam.ac.uk).

Axonal site of spike initiation enhances auditory coincidence detection

Hiroshi Kuba¹, Takahiro M. Ishii¹ & Harunori Ohmori¹

Neurons initiate spikes in the axon initial segment or at the first node in the axon^{1–4}. However, it is not yet understood how the site of spike initiation affects neuronal activity and function. In nucleus laminaris of birds, neurons behave as coincidence detectors for sound source localization and encode interaural time differences (ITDs) separately at each characteristic frequency (CF)^{5–7}. Here we show, in nucleus laminaris of the chick, that the site of spike initiation in the axon is arranged at a distance from the soma, so as to achieve the highest ITD sensitivity at each CF. Na⁺ channels were not found in the soma of high-CF (2.5–3.3 kHz) and middle-CF (1.0–2.5 kHz) neurons but were clustered within a short segment of the axon separated by 20–50 μ m from the soma; in low-CF (0.4–1.0 kHz) neurons they were clustered in a longer stretch of the axon closer to the soma. Thus, neurons initiate spikes at a more remote site as the CF of neurons increases. Consequently, the somatic amplitudes of both orthodromic and antidromic spikes were small in high-CF and middle-CF neurons and were large in low-CF neurons. Computer simulation showed that the geometry of the initiation site was optimized to reduce the threshold of spike generation and to increase the ITD sensitivity at each CF. Especially in high-CF neurons, a distant localization of the spike initiation site improved the ITD sensitivity because of electrical isolation of the initiation site from the soma and dendrites, and because of reduction of Na⁺-channel inactivation by attenuating the temporal summation of synaptic potentials through the low-pass filtering along the axon.

ITD is an important cue for localizing sound sources in the horizontal plane⁶. Nucleus laminaris (NL) neurons of birds detect the coincidence of binaural synaptic inputs and calculate ITDs in the microsecond range⁷. Recent works have revealed that several mechanisms underlie this coincidence detection, including the following: first, acceleration of excitatory postsynaptic potential time course by K⁺ channel activation^{8,9}; second, inhibitory synaptic inputs^{10,11}; third, segregation of binaural inputs by dendrites¹²; and fourth, depression of synaptic transmission^{13,14}. NL neurons are arranged tonotopically in a rostro-medial (high CF) to caudo-lateral (low CF) direction within the nucleus¹⁵ (Supplementary Fig. 1a), and show several CF-dependent specializations such as morphology of dendrites¹⁶ and K⁺-channel expression¹⁷. Axon hillock and initial segment are myelinated in the high-CF neuron, implying that spikes may be generated at a different site of neurons depending on the CF¹⁸. Here we examined the site of action potential initiation in NL neurons along the tonotopic axis and explored how the location contributes to the coincidence detection.

In the high-CF and middle-CF neurons, both orthodromic (thick traces in Fig. 1a) and antidromic (thin traces) spikes were smaller than those in low-CF neurons. Moreover, the antidromic spike of the high-CF and middle-CF neurons was far smaller than the orthodromic one ($P < 0.01$, $n = 11$ cells), although the maximum rate of

rise of these two spikes was not different (see Supplementary Fig. 1b). The small antidromic spike in the high-CF and middle-CF neurons was not an abortive response for the following reasons. First, although antidromic spikes during somatic subthreshold depolarization exceeded the threshold of orthodromic spikes, further regenerative responses were not induced (Fig. 1b, left, and insets for expanded time course, observed in eight cells). Second, the antidromic and the orthodromic spikes blocked each other when evoked at time intervals shorter than 0.5 ms (Fig. 1c, filled circles, seven cells),

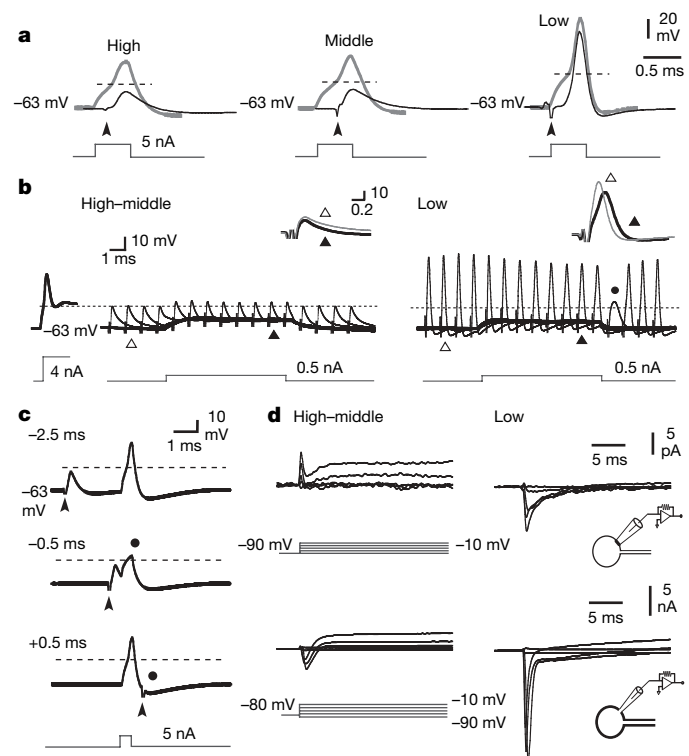


Figure 1 | The site of action potential generation is remote from the cell soma in high-CF and middle-CF neurons. **a**, Antidromic (thin black lines) and orthodromic (thick grey lines) spikes at each CF. Left, high CF; middle, middle CF; right, low CF. Broken lines indicate threshold. Orthodromic spikes were elicited by somatic current injection (lower traces), and antidromic spikes by stimulating the axon (arrowheads). **b**, Antidromic spikes during somatic depolarization. Left, high CF and middle CF; right, low CF. Inset, spikes before and during depolarization recorded at the corresponding symbols below. **c**, Antidromic spike paired with orthodromic spike; high CF and middle CF. **d**, Voltage-clamp recordings under cell-attached (upper) and whole-cell (lower) conditions. Left, high CF and middle CF; right, low CF.

¹Department of Physiology, Faculty of Medicine, Kyoto University, Kyoto, 606-8501, Japan.

suggesting that both spikes have the same origin. These observations indicate that spikes are generated at a remote site from the soma in the high-CF and middle-CF neurons. In contrast, in the low-CF neuron, antidromic spikes were affected and eventually interrupted by the somatic depolarization (Fig. 1b, right, filled circle, and insets, five cells), indicating that spikes are generated at a closer site.

Localization of voltage-gated Na^+ (Nav) channels was studied electrophysiologically (Fig. 1d). In the low-CF neurons, cell-attached patch recordings from the soma occasionally showed Na^+ currents (5 of 22 cells, 13.3 ± 1.4 pA at -30 mV, ranging from 10.7 to 18.1 pA; upper right), but not in the high-CF and middle-CF neurons (26 cells; upper left). Under the whole-cell recording condition, depolarizing steps evoked rapidly inactivating tetrodotoxin-sensitive Na^+ currents in both high-CF and middle-CF neurons and low-CF neurons, and the amplitude at -30 mV was fivefold to sixfold smaller in the high-CF and middle-CF neurons (2.9 ± 0.2 nA, $n = 8$ cells; lower left) than in the low-CF neuron (16.8 ± 1.4 nA, $n = 9$ cells; lower right). These observations may indicate that Nav channels are absent from the cell soma in the high-CF and middle-CF neurons. Even in the low-CF neurons, Nav channel density in the soma was small; the Na^+ current in the patch (0.67 ± 0.07 pA μm^{-2} , $n = 5$ cells, $20 \mu\text{m}^2$;

see Supplementary Methods) was about 30-fold smaller than that expected from the somatic surface area (21.0 ± 1.8 pA μm^{-2} , $n = 9$ cells, $800 \mu\text{m}^2$)¹⁷.

These electrophysiological observations were consistent with the following immunohistological localization of Nav channels (Fig. 2). An antibody for the neuronal Nav channels (Pan-Nav) showed fibrous clusters (hot spots; red arrowheads in Fig. 2a) and many puncta presumably representing the nodes of Ranvier. The length of the hot spot increased towards the low-CF region (Fig. 2g). The hot spot was also immunopositive against Nav1.6 (Fig. 2b), which is known to be expressed in the myelinated axon in the brain¹⁹. Nav channels were also co-localized with voltage-gated K^+ channels (Kv1.2) (Fig. 2c). Thus, the hot spot may correspond to the axon initial segment rather than the node of Ranvier, because Kv channels are confined to perinodal regions at the node²⁰. Subcellular localization of the hot spot was examined by immunostaining after the identification of NL neurons with a retrograde tracer. Nav channels were labelled in the axon in both high-CF and middle-CF neurons and low-CF neurons (Fig. 2d, e, and Supplementary Fig. 2). In the high-CF and middle-CF neurons, the hot spot started at a more distal site from the soma (the separation of the hot spot from the soma is defined as distance, Fig. 2f) and was shorter in length than in the low-CF neurons (the stretch of the hot spot is defined as length, Fig. 2g). The length of hot spot had a clear negative correlation with the distance (Fig. 2h).

How these geometries of the hot spot affect firing efficacy and coincidence detection of a neuron was examined by simulation with a multiple-compartment model. For simplicity, the geometry of soma-dendrites was assumed to be independent of CF^{16,17}, whereas the lengths of the hot spot (length, L) and myelinated initial segment (distance, D) were varied depending on the frequency region of the model neuron. The model well reproduced the action potentials observed at each CF when the corresponding distance and length of hot spot were assigned (Fig. 3a, b, and Supplementary Figs 3 and 4). The excitability of the model neuron was evaluated as the threshold current (0.3 ms in duration) required for generating an action potential. The geometry of the hot spot critically affected the threshold current (Fig. 3c, and Supplementary Fig. 5). When the hot spot was short (10 μm in length; arrow f in Fig. 3c), increasing the hot spot distance reduced the threshold current, and the minimum (red, indicating the highest excitability) was found at a distance of 20–30 μm . With a long hot spot (25 μm in length; arrow h in Fig. 3c), however, the minimum was found with a shorter distance (0 μm). The distribution of threshold current almost overlapped with the observed distribution of the hot spot in the NL neuron (dots in Fig. 3c); the observed distribution was slightly more distant in the high-CF neurons (30–80 μm ; white dots). Thus, the distance and the length of hot spot were arranged to increase the excitability of NL neurons. Spike generation was significantly affected by the somatic capacitance and K^+ current. Therefore, either the isolation of the hot spot from the soma or an increase in the hot spot length was required to reduce the threshold current. However, a further increase in the distance or the length of hot spot increased the threshold current, because of the dissipation of charging current across the axonal membrane or the increased axonal K^+ current (see Supplementary Figs 5 and 6).

Next we tested how the geometry of the hot spot affects the ITD sensitivity by applying trains of binaural synaptic inputs into the soma at various phase differences (Fig. 3d, boxed insets; 180° corresponds to 0.17 ms of ITDs at 3 kHz CF; see Supplementary Methods). The model neuron could respond to ITDs at the maximum firing rate for the synaptic inputs applied in phase (0°) and at the minimum rate for the out-of-phase (180°) inputs (Fig. 3d, e). The spikes were phase-locked to the in-phase inputs (standard deviation of peak times, 41 μs). The firing rate and the ITD sensitivity, defined here as the difference of firing rate between in-phase and out-of-phase inputs, increased with the increase in input conductance (Fig. 3e), as was observed in the barn-owl NL neurons *in vivo*¹⁰.

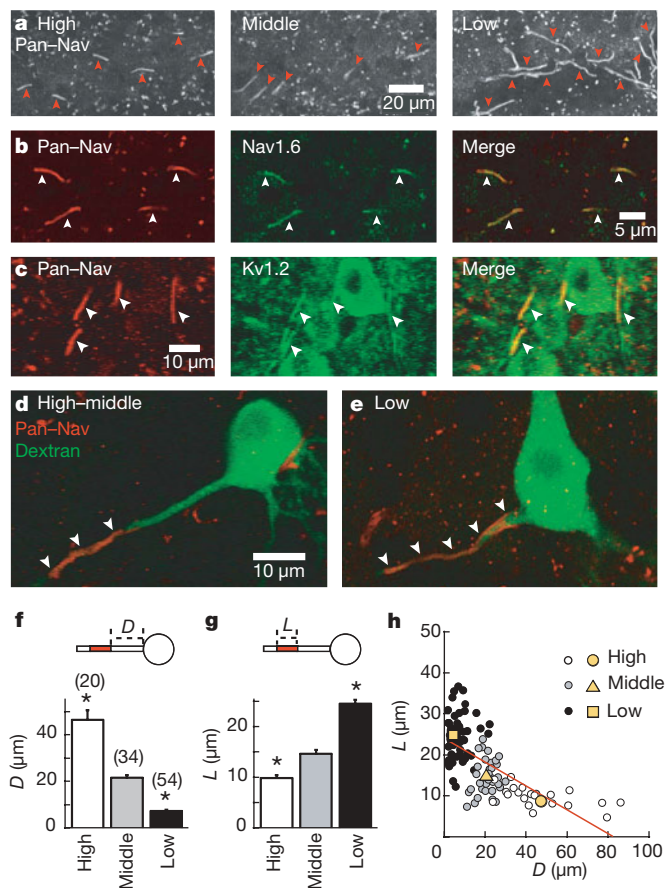


Figure 2 | Nav channels cluster in the axon at some distance from the soma in high-CF and middle-CF neurons. **a**, Fibrous representation of Nav (hot spot, arrowheads) in NL. Left, high CF; middle, middle CF; right, low CF. **b**, Double immunostaining with Pan-Nav (red, left) and Nav1.6 (green, middle). Right, merged images. **c**, Colocalization of Nav (red, left) with Kv1.2 (green, middle). Right, merged images. **d**, **e**, Nav channels (red) are present on an axon of retrogradely labelled NL neurons (green) in both high-CF and middle-CF neurons (**d**) and low-CF neurons (**e**). **f**, **g**, Distance D (**f**) and length L (**g**) of Nav hot spot. Error bars indicate standard errors. Asterisk, $P < 0.01$ compared with middle CF. Numbers in parentheses are the number of cells. **h**, Negative correlation between length and distance: $y = -0.29x + 24.0$, $r = 0.66$, $P < 0.01$, $n = 108$ cells from **f** and **g**. Yellow symbols are the averages.

ITD sensitivity was calculated as a function of the distance of the hot spot, for model neurons having the hot spot of a corresponding length at each frequency input (Fig. 3f–h; see Fig. 2h): Supplementary Fig. 7 complementarily shows the effects of hot spot length on the ITD sensitivity. The input conductance was set to fire the neuron at about 600 Hz (ref. 10) (Supplementary Methods; see effects of input conductance on ITD sensitivity in Supplementary Fig. 8). The ITD sensitivity depended on the distance of the hot spot from the soma (Fig. 3f–h); the maximum (arrowheads) was given at 0–10 μm for the low CF, 20–40 μm for the middle CF, and 50–60 μm for the high CF. This tendency was the same as that observed in the immunohistochemistry (yellow symbols and red broken lines; see Fig. 2h) and the distribution of threshold current (lower panels of Fig. 3f–h; see Fig. 3c). However, there was a dissociation in the distance between the maximum ITD sensitivity and the minimum threshold current (asterisks and black broken lines), particularly at high-CF neurons (Fig. 3f); the maximum ITD sensitivity was given at a more remote site.

To explore the causes of this difference, we examined the effects of a high-frequency burst of synaptic inputs on the membrane potential at the hot spot, under the block of spike generation (Fig. 4a, thick black traces). During high-frequency inputs (3 kHz), synaptic potentials were temporally summated to depolarize the hot spot located near the soma by about 20 mV (Fig. 4a, top trace). This depolarization strongly inactivated Na^+ channels and reduced the excitability

(Fig. 4b, open triangle). However, when the hot spot was separated from the soma (50 μm distance), the distance attenuated the depolarization (Fig. 4a, bottom trace) and reduced the extent of inactivation of Na^+ channels at the hot spot (250% increase in h_{Na} ; Fig. 4b, filled triangle; Fig. 4c, bottom). The distance also decreased the activation of K^+ current at the hot spot, but the effect was small (less than 25%). As a result, when neurons receive high-frequency synaptic inputs (3 kHz), the hot spot to give the highest excitability (minimum threshold current) was shifted towards a distant location (from D 20 μm without inputs to D 40 μm with inputs; Fig. 4c, middle). Thus, the remote localization of the hot spot was preferable for the detection of ITDs, when a high-frequency burst of inputs arrived (see frequency effects in Supplementary Fig. 9).

We have shown here that the reduction of Na^+ channel inactivation is crucial in detecting ITDs at high sensitivity during high-frequency inputs. The findings are consistent with the observation that the initial segment is myelinated in high-frequency NL neurons of the barn owl¹⁸. In the medial superior olive, the mammalian homologue of NL, the spike amplitude was small and showed a short latency when recorded near the axon in gerbils²¹, indicating that axonal spike initiation is probably a common feature of binaural coincidence detectors in birds and mammals. Moreover, recent works have suggested that other central neurons initiate spikes at a more distant site than the axon initial segment^{2,3,22}. Although the functional significance of axonal spike initiation in those neurons remains unknown,

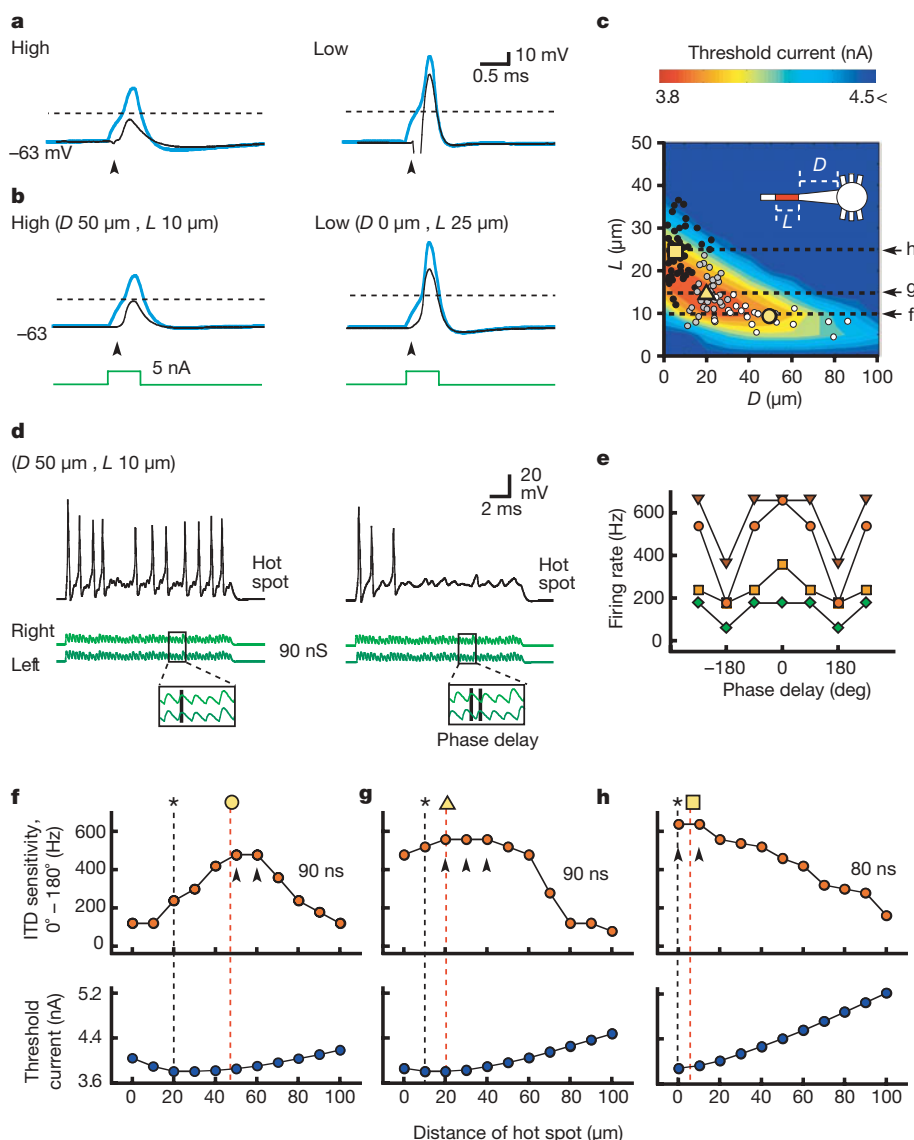


Figure 3 | Nav channel distribution is optimized for ITD detection.

a, b, Recorded (**a**) and simulated (**b**) action potentials. Broken lines indicate threshold. Thick blue lines, orthodromic; thin black lines, antidromic. **c,** Threshold current (see the text) as a function of length L and distance D . Dots and yellow symbols are from Fig. 2h. The notation '4.5<' stands for injected current of larger than 4.5 nA. **d,** Responses to binaural synaptic inputs of 3 kHz. The binaural phase delay was 0° (left) and 180° (right). **e,** Firing rate as a function of phase and input conductance (triangles, 98 nS; circles, 90 nS; squares, 82 nS; diamonds, 74 nS). **f–h,** ITD sensitivity (top) and threshold current (bottom) as a function of hot spot distance. **f,** High CF (L 10 μm ; 3 kHz); **g,** middle CF (L 15 μm ; 2 kHz); **h,** low CF (L 25 μm ; 1 kHz). Asterisks and black broken lines indicate the threshold minimum calculated in **c**. Yellow symbols and red broken lines are the observed average distances from Fig. 2h. Arrowheads indicate the maximum ITD sensitivity.

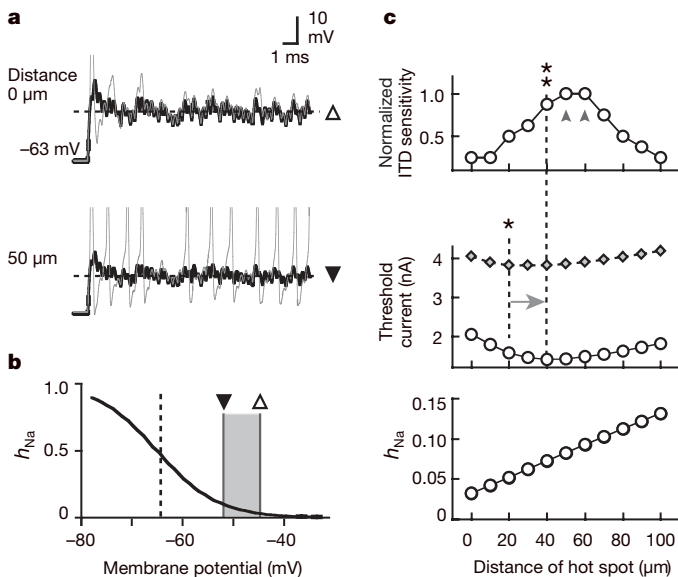


Figure 4 | Cable properties of axon enhance ITD sensitivity during high-frequency inputs. **a**, Membrane potentials of hot spot located at 0-μm (upper) and 50-μm (lower) distance during binaural 3-kHz inputs (0°, 90 nS) in high-CF neuron (10-μm length). Plateau depolarization was defined as an average of membrane potential (horizontal broken lines and triangles). **b**, Voltage dependence of h_{Na} (inactivation parameter) during depolarization. Filled triangle, 50 μm with inputs; open triangle, 0 μm with inputs. Vertical broken line indicates the rest. **c**, ITD sensitivity (top, from Fig. 3f), threshold current (middle) and h_{Na} (bottom) as a function of hot spot distance. In the middle panel, circles represent with inputs, and diamonds represent without inputs. The plateau depolarization (induced by constant current inputs to soma, 5.5 nA) displaced the threshold minimum to a distant location (middle, with inputs, double asterisk). The single asterisk indicates threshold minimum without inputs.

the present results help our understanding of how the site of action potential generation can fine-tune neuronal computations.

METHODS

Whole-cell or cell-attached recordings were made from NL neurons in brainstem slices (200–250 μm) of chicks at 3–7 days after hatching (Fig. 1)²³. Antibodies against Pan-Nav and Nav1.6 were raised for immunohistochemistry (Fig. 2)^{17,24}. Some immunostainings were made after revealing NL neurons with a retrograde tracer. Neuronal modelling and simulation were performed with NEURON 5.6 (Figs 3 and 4). K^+ and Na^+ currents in the model followed the web-accessible Model DB^{25–28}. Statistical significance was tested with Student's *t*-test. Values are presented as means ± s.e.m. Detailed methods are provided in Supplementary Methods.

Received 5 August; accepted 13 October 2006.

Published online 29 November 2006.

1. Stuart, G., Spruston, N., Sakmann, B. & Häusser, M. Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends Neurosci.* **20**, 125–131 (1997).
2. Clark, B. A., Monsivais, P., Branco, T., London, M. & Häusser, M. The site of action potential initiation in cerebellar Purkinje neurons. *Nature Neurosci.* **8**, 137–139 (2005).
3. Palmer, L. M. & Stuart, G. J. Site of action potential initiation in layer 5 pyramidal neurons. *J. Neurosci.* **26**, 1854–1863 (2006).
4. Khaliq, Z. M. & Raman, I. M. Relative contribution of axonal and somatic Na channels to action potential initiation in cerebellar Purkinje neurons. *J. Neurosci.* **26**, 1935–1944 (2006).
5. Knudsen, E. I. & Konishi, M. Mechanisms of sound localization in the barn owl (*Tyto alba*). *J. Comp. Physiol. [A]* **133**, 13–21 (1979).

6. Moiseff, A. & Konishi, M. Neuronal and behavioral sensitivity to binaural time differences in the owl. *J. Neurosci.* **1**, 40–48 (1981).
7. Carr, C. E. & Konishi, M. A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci.* **10**, 3227–3246 (1990).
8. Reyes, A. D., Rubel, E. W. & Spain, W. J. *In vitro* analysis of optimal stimuli for phase-locking and time-delayed modulation of firing in avian nucleus laminaris neurons. *J. Neurosci.* **16**, 993–1007 (1996).
9. Kuba, H., Yamada, R. & Ohmori, H. Evaluation of the limiting acuity of coincidence detection in nucleus laminaris of the chicken. *J. Physiol. (Lond.)* **552**, 611–620 (2003).
10. Pena, J. L., Viète, S., Albeck, Y. & Konishi, M. Tolerance to sound intensity of binaural coincidence detection in the nucleus laminaris of the owl. *J. Neurosci.* **16**, 7046–7054 (1996).
11. Funabiki, K., Koyano, K. & Ohmori, H. The role of GABAergic inputs for coincidence detection in the neurons of nucleus laminaris of the chick. *J. Physiol. (Lond.)* **508**, 851–869 (1998).
12. Agmon-Snir, H., Carr, C. E. & Rinzel, J. The role of dendrites in auditory coincidence detection. *Nature* **393**, 268–272 (1998).
13. Kuba, H., Koyano, K. & Ohmori, H. Synaptic depression improves coincidence detection in the nucleus laminaris in brainstem slices of the chick embryo. *Eur. J. Neurosci.* **15**, 984–990 (2002).
14. Cook, D. L., Schwindt, P. C., Grande, L. A. & Spain, W. J. Synaptic depression in the localization of sound. *Nature* **421**, 66–70 (2003).
15. Rubel, E. W. & Parks, T. N. Organization and development of the brain stem auditory nuclei of the chicken: tonotopic organization of N. magnocellularis and N. laminaris. *J. Comp. Neurol.* **164**, 411–434 (1975).
16. Smith, D. J. & Rubel, E. W. Organization and development of brain stem auditory nuclei of the chicken: dendritic gradients in nucleus laminaris. *J. Comp. Neurol.* **186**, 213–240 (1979).
17. Kuba, H., Yamada, R., Fukui, I. & Ohmori, H. Tonotopic specialization of auditory coincidence detection in nucleus laminaris of the chick. *J. Neurosci.* **25**, 1924–1934 (2005).
18. Carr, C. E. & Boudreau, R. E. An axon with a myelinated initial segment in the bird auditory system. *Brain Res.* **628**, 330–334 (1993).
19. Caldwell, J. H., Schaller, K. L., Lasher, R. S., Peles, E. & Levinson, S. R. Sodium channel Nav1.6 is localized at nodes of Ranvier, dendrites, and synapses. *Proc. Natl Acad. Sci. USA* **97**, 5616–5620 (2000).
20. Rasband, M. N. & Shrager, P. Ion channel sequestration in central nervous system axon. *J. Physiol. (Lond.)* **525**, 63–73 (2000).
21. Scott, L. L., Mathews, P. J. & Golding, N. L. Posthearing developmental refinement of temporal processing in principal neurons of the medial superior olive. *J. Neurosci.* **25**, 7887–7895 (2005).
22. Colbert, C. M. & Johnston, D. Axonal action-potential initiation and Na^+ channel densities in the soma and axon initial segment of subicular pyramidal neurons. *J. Neurosci.* **16**, 6676–6686 (1996).
23. Kuba, H., Koyano, K. & Ohmori, H. Development of membrane conductance improves coincidence detection in the nucleus laminaris of the chicken. *J. Physiol. (Lond.)* **540**, 529–542 (2002).
24. Yamada, R., Kuba, H., Ishii, T. M. & Ohmori, H. Hyperpolarization-activated cyclic nucleotide-gated cation channels regulate auditory coincidence detection in nucleus laminaris of the chick. *J. Neurosci.* **25**, 8867–8877 (2005).
25. Rathouz, M. & Trussell, L. Characterization of outward currents in neurons of the avian nucleus magnocellularis. *J. Neurophysiol.* **80**, 2824–2835 (1998).
26. Simon, J. Z., Carr, C. E. & Shamma, S. A. A dendritic model of coincidence detection in the avian brainstem. *Neurocomputing* **26–27**, 263–269 (1999).
27. Rothman, J. S. & Manis, P. B. The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons. *J. Neurophysiol.* **89**, 3097–3113 (2003).
28. Hines, M. L., Morse, T., Migliore, M., Carnevale, N. T. & Shepherd, G. M. Model DB: a database to support computational neuroscience. *J. Comput. Neurosci.* **17**, 7–11 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. O. Trussell for reading the manuscript, T. Kaneko for advice on preparing antibodies, and M. Dezawa for instruction on immunostaining. This work was supported by a Grant-in-aid from MEXT to H.K. and H.O.

Author Contributions H.K. performed all experiments and simulations. T.M.I. contributed to preparing antibodies. H.K. and H.O. wrote the paper together.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.O. (ohmori@nbiol.med.kyoto-u.ac.jp).

Early events in the thymus affect the balance of effector and regulatory T cells

Daniel J. Pennington^{1,2}, Bruno Silva-Santos^{1,3}, Tobias Silberzahn¹, Mónica Escórcio-Correia¹, Martin J. Woodward⁴, Scott J. Roberts⁵, Adrian L. Smith⁶, P. Julian Dyson⁷ & Adrian C. Hayday¹

In cellular immunology the critical balance between effector and regulatory mechanisms is highlighted by serious immunopathologies attributable to mutations in *Foxp3*, a transcription factor required for a major subset of regulatory T (T_R) cells^{1–3}. Thus, many studies have focused on the developmental origin of T_R cells, with the prevailing view that they emerge in the thymus from late-stage T-cell progenitors whose T-cell receptors (TCRs) engage high affinity (agonist) ligands^{4–6}. This study questions the completeness of that interpretation. Here we show that without any obvious effect on TCR-mediated selection, the normal differentiation of mouse $\gamma\delta$ T cells into potent cytolytic and interferon- γ -secreting effector cells is switched towards an aggregate regulatory phenotype by limiting the capacity of $CD4^+CD8^+$ T-cell progenitors to influence *in trans* early $\gamma\delta$ cell progenitors. Unexpectedly, we found that the propensity of early TCR- $\alpha\beta^+$ progenitors to differentiate into *Foxp3*⁺ T_R cells is also regulated *in trans* by $CD4^+CD8^+$ T-cell progenitor cells, before agonist selection.

Thymocyte differentiation progresses from early $CD4^-CD8^-$ (double negative) progenitors either directly to TCR- $\gamma\delta^+$ cells, or to $CD4^+CD8^+$ (double positive) cells, which are selected by TCR engagement into mature 'single-positive' $CD4^+$ and $CD8^+$ TCR- $\alpha\beta^+$ cells. The double-negative to double-positive transition depends on a pre-TCR comprising TCR- β and pre-T α chains, and mice lacking either harbour very few double-positive cells. Of note, double-positive cells function not only as progenitors of single-positive cells, but also 'condition' double-negative cell differentiation *in trans*⁷. Reflecting this, $\gamma\delta$ cells in pT α -deficient (*Ptcr α* ^{-/-}) and TCR- β -deficient (*Tcrb*^{-/-}) mice differentiate with abnormally low expression of interferon- γ (IFN- γ) and cytolytic effector molecules. Nonetheless, other gene products (for example, XCL1) are expressed normally⁷, suggesting that rather than being generally impaired, $\gamma\delta$ cell progenitors lacking *trans*-conditioning might differentiate along an alternative pathway. Consistent with this, we found that genes expressed by activated splenic $\gamma\delta$ cells in *Tcrb*^{-/-} but not wild-type mice resembled those expressed by activated skin-resident dendritic epidermal T cells (DETCs) that naturally develop in the fetus before double-positive cells accumulate⁸ (Fig. 1a). Resting splenic $\gamma\delta$ cells showed less resemblance (Fig. 1b), which, together with a lack of DETC-associated V γ 5 expression (Supplementary Fig. 1), refutes the possibility that *Tcrb*^{-/-} splenocytes are simply DETCs aberrantly localized to the spleen.

On several genetic backgrounds, DETC-deficient mice spontaneously develop cutaneous immunopathologies attributable to dysregulated $\alpha\beta$ T cells⁹. Thus, although DETCs display cytolytic and pro-inflammatory potentials¹⁰, their aggregate phenotype *in vivo* is

regulatory. We therefore determined whether activated splenic $\gamma\delta$ cells in *Tcrb*^{-/-} mice also display aggregate regulatory activity. The membranes of TCR- $\alpha\beta^+CD25^{lo}CD4^+$ effector cells were labelled with carboxyfluorescein diacetate succinimidyl ester (CFSE), the dilution of which monitors cytokinesis. At 72 h or 86 h after activation only ~5–10% of cells remained undivided (Fig. 1c). Conversely, in 1:1 and 1:0.1 co-cultures with unlabelled $CD25^{hi}CD4^+$ T_R cells^{11,12}, ~50% and ~20%, respectively, of cells remained undivided. 1:1 co-cultures with wild-type $\gamma\delta$ splenocytes conferred no such regulation, whereas in co-cultures with $\gamma\delta$ cells from *Tcrb*^{-/-} mice, ~25% of effector T cells remained undivided. This suppressive capacity did not require $\gamma\delta$ cell subsetting, for example, by CD25 expression. Thus, $\gamma\delta$ splenocytes developing independently of a normal double-positive pool resemble activated DETCs: both populations contain cytolytic and IFN- γ -producing effectors, but on aggregate are regulatory.

Immunoregulation is a critically important function provided by different mechanisms. Whereas natural $\alpha\beta$ T_R cell activity relies on *Foxp3* transcription^{1–3}, other regulatory cells, including DETCs, do not¹⁰. Indeed, *Foxp3* has not been reported in $\gamma\delta$ cells, and we could not detect *Foxp3* RNA in TCR- $\gamma\delta^+$ splenocytes from C57.BL/6J or FVB/n mice, or from TCR- α -deficient (*Tcr α* ^{-/-}) mice that lack $\alpha\beta$ T cells but harbour a large double-positive compartment (Fig. 2a). By contrast, *Foxp3* RNA was amplified from $\gamma\delta$ cells of *Ptcr α* ^{-/-} and *Tcrb*^{-/-} mice, and *Foxp3*-protein-expressing $\gamma\delta$ cells were reproducibly detected among splenocytes from *Tcrb*^{-/-} but not wild-type mice (Fig. 2a, b). Similarly, TCR- $\gamma\delta^+$ double-negative thymocytes from *Tcrb*^{-/-} and *Ptcr α* ^{-/-} (but not wild type) mice also expressed *Foxp3* (Fig. 2c, d). Thus, although it was only in a minority (1–2%) of TCR- $\gamma\delta^+$ cells (Supplementary Fig. 2), the expression of *Foxp3* in *Tcrb*^{-/-} mice emphasized the switch towards aggregate regulatory T-cell differentiation in the context of double-positive deficiency. Indeed, *Foxp3* RNA was also detected in $\gamma\delta$ thymocytes from mice lacking lymphotoxin- β receptor (LT β R) (Fig. 2d) that contributes to *trans*-conditioning by double-positive cells⁷.

Many experiments support the proposal that $CD4^+CD25^{hi}$ TCR- $\alpha\beta^+$ T_R cells are selected by high-affinity TCR agonist engagement of double-positive cells^{4–6}. Thus, *Foxp3*⁺ $\gamma\delta$ cells in *Tcrb*^{-/-} mice might reflect an artefactual rescue by TCR- $\gamma\delta$ of double-positive cells that then engage agonist and select as T_R cells. However, whereas TCR- $\gamma\delta$ will rescue some double-positive differentiation in *Tcrb*^{-/-} mice¹³, such TCR- $\gamma\delta^+$ double-positive cells usually die without maturation^{13–15}. Indeed, *Foxp3* RNA was not detectable in double-positive cells from *Tcrb*^{-/-} mice (Fig. 2e). Moreover, the 'rescue hypothesis' does not obviously explain *Foxp3* RNA in LT β R-deficient (*Ltbr*^{-/-})

¹Peter Gorer Department of Immunobiology, King's College London School of Medicine, Guy's Hospital, London SE1 9RT, UK. ²Institute of Cell and Molecular Science, The London and Barts School of Medicine and Dentistry, Queen Mary College, London E1 2AT, UK. ³Instituto de Medicina Molecular, Faculdade de Medicina de Lisboa, Lisboa 1649-028, Portugal. ⁴Molecular Haematology and Cancer Biology Unit, Camelia Botnar Laboratories, Institute of Child Health, University College, London WC1N 1EH, UK. ⁵Department of Dermatology, 15 York Street HRT 603, Yale University School of Medicine, New Haven, Connecticut 06510, USA. ⁶Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berks RG20 7NN, UK. ⁷Division of Medicine, Imperial College, Hammersmith Hospital, Du Cane Road, London W12 0HS, UK.

$\gamma\delta$ thymocytes (Fig. 2d). Instead, we pursued the hypothesis that the aggregate regulatory phenotype of $\gamma\delta$ cells in *Tcrb*^{-/-} mice might reflect an uncharacterized route of regulatory T-cell differentiation.

Despite pT α deficiency, some double-negative to double-positive maturation accompanied by limited cell expansion occurs in *Ptcr*^{-/-} mice, giving rise to a TCR- $\alpha\beta$ ⁺CD4⁺ thymocyte pool ~130 times smaller than wild type¹⁶ (Supplementary Table 1). CD25^{hi}Foxp3⁺ cell numbers are similarly reduced, confirming that the pre-TCR is necessary for T_R-cell differentiation, but this decline is only ~20-fold, and 25–30% of *Ptcr*^{-/-} CD4⁺ thymocytes are CD25^{hi}Foxp3⁺ compared with 2.5–4.5% in wild-type mice (Fig. 3a). Hence, effector and regulatory CD4⁺ T cells do not share the same dependence on pT α . CD4⁺ cell numbers in the periphery

of *Ptcr*^{-/-} mice were less overtly reduced (possibly because of homeostatic expansion), but again effects on T_R cells were more mild, with the CD25^{hi}Foxp3⁺ phenotype shown by ~23–32% of CD4⁺ splenocytes compared with 6–11% in wild type (Fig. 3b; see also Supplementary Table 1). Moreover, regulatory function per cell was equal to or greater than wild type (Fig. 3c). This enrichment for T_R cells in *Ptcr*^{-/-} mice is highly selective: for example, intestinal TCR- $\alpha\beta$ ⁺CD8 α ⁺ cells are markedly depleted (Supplementary Table 1). Because pT α function precedes double-positive differentiation, its mutation would not be predicted to affect agonist selection into mature CD4⁺ (SP4) thymocytes, supporting which the CD4⁺ thymocyte TCR repertoire is equivalent to wild type (Supplementary Table 2).

To verify independently the relationship between pT α , double-positive cells and T_R differentiation, we examined pT α transgenic

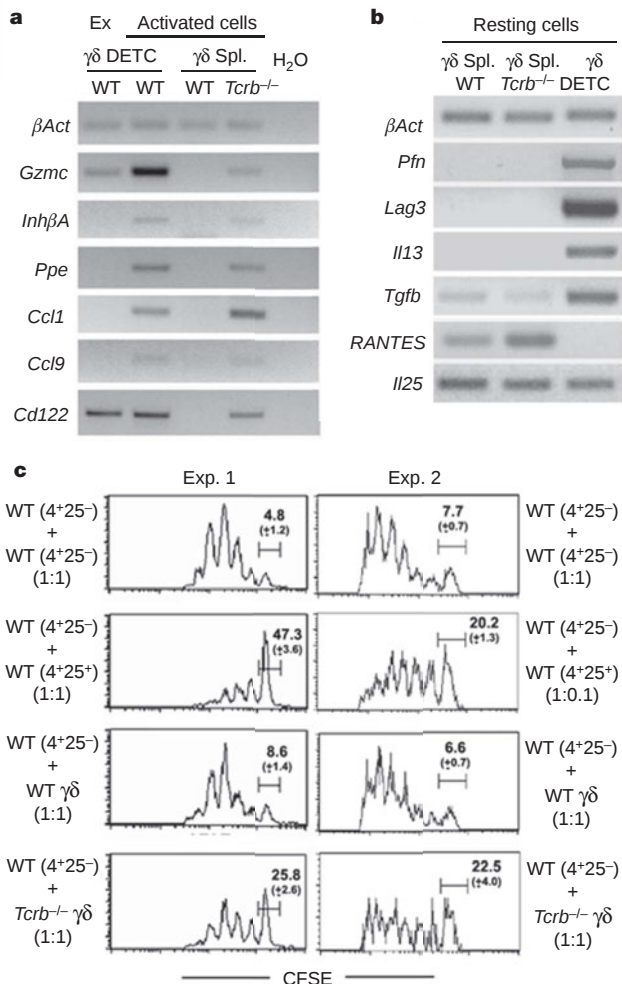


Figure 1 Activated splenic $\gamma\delta$ cells from *Tcrb*^{-/-} mice express activated DETC-specific genes and display regulatory potential *in vitro*. **a**, Semi-quantitative RT-PCR for genes as indicated, on representative ($n > 3$) cDNA from wild-type TCR- $\gamma\delta$ ⁺ DETCs *ex vivo* (Ex, $\gamma\delta$ DETC, WT), and anti-CD3 ϵ monoclonal-antibody-activated wild-type TCR- $\gamma\delta$ ⁺ DETCs (activated cells, $\gamma\delta$ DETC, WT), wild-type splenic TCR- $\gamma\delta$ ⁺ cells (activated cells, $\gamma\delta$ Spl., WT), and *Tcrb*^{-/-} splenic TCR- $\gamma\delta$ ⁺ cells (activated cells, $\gamma\delta$ Spl., *Tcrb*^{-/-}). **b**, Semi-quantitative RT-PCR for genes as indicated, on representative ($n > 3$) cDNA from resting wild-type splenic TCR- $\gamma\delta$ ⁺ cells ($\gamma\delta$ Spl., WT), resting *Tcrb*^{-/-} splenic TCR- $\gamma\delta$ ⁺ cells ($\gamma\delta$ Spl., *Tcrb*^{-/-}) or resting *ex vivo* wild-type TCR- $\gamma\delta$ ⁺ DETCs ($\gamma\delta$, DETC). **c**, CFSE-labelled wild-type CD4⁺CD25⁻ splenic T cells (5×10^4) were cultured (in triplicate) with either wild-type splenic CD4⁺CD25⁻ cells (5×10^4), wild-type splenic CD4⁺CD25⁺ cells (5×10^4 (Exp. 1); 5×10^3 (Exp. 2)), and splenic $\gamma\delta$ cells (5×10^4) from wild-type or *Tcrb*^{-/-} mice, plus soluble anti-CD3 ϵ monoclonal antibody (+antigen presenting cells) for 72 h (Exp. 1) or 86 h (Exp. 2). Average percentage of undivided cells ($\pm$ s.d.) is indicated.

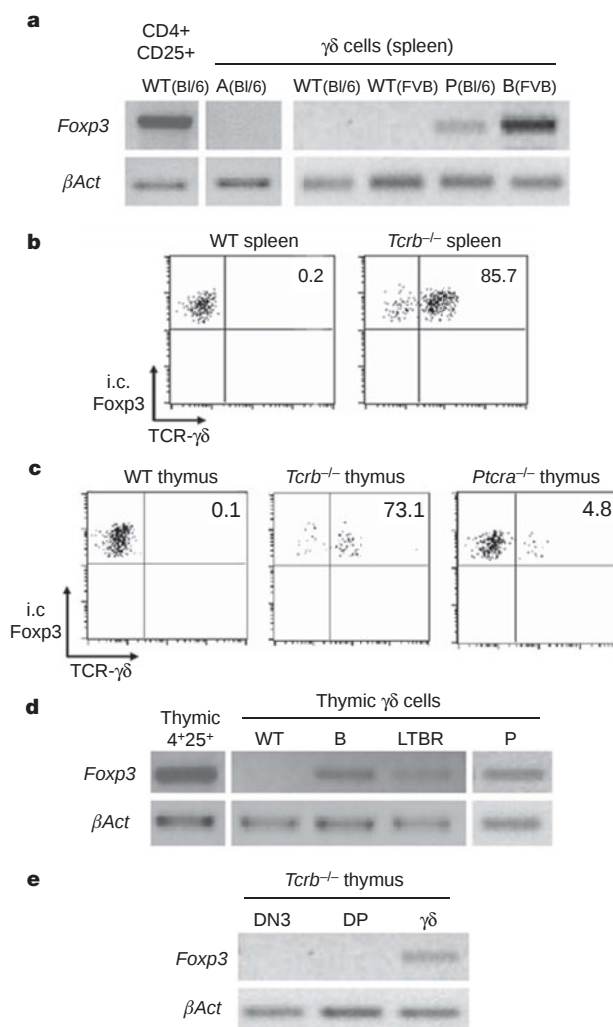


Figure 2 $\gamma\delta$ cells from *Tcrb*^{-/-} and *Ptcr*^{-/-} mice express Foxp3. **a**, Representative ($n > 3$) RT-PCR for Foxp3 (compared to β -actin) in splenic $\gamma\delta$ cells from C57Bl/6J or FVB/n strains of wild-type (WT), *Tcr*^{-/-} (A), *Ptcr*^{-/-} (P) or *Tcrb*^{-/-} (B) mice. The Foxp3⁺ control is wild-type splenic TCR- $\alpha\beta$ ⁺CD4⁺CD25⁺ cells. **b**, **c**, Representative ($n > 6$) intracellular (i.c.) Foxp3 against TCR- $\gamma\delta$ plots of (Foxp3⁺) splenocytes from wild-type and *Tcrb*^{-/-} mice, or thymocytes from wild-type, *Tcrb*^{-/-} or *Ptcr*^{-/-} mice. As these gatings represent minor populations (see text), representative un-gated plots are shown in Supplementary Fig. 2. Quadrant percentages are indicated. **d**, **e**, Representative ($n > 3$) RT-PCR for Foxp3 in thymocytes from wild-type (WT), *Tcrb*^{-/-} (B), *Ltbr*^{-/-} (LTBR) or *Ptcr*^{-/-} (P) mice as in **a**. DN3 is CD4⁻CD8 α ⁻TCR- $\alpha\beta$ ⁻TCR- $\gamma\delta$ ⁻B220⁻NK1.1⁻DX5⁻CD25⁺CD44⁻; DP (double positive) is CD4⁺CD8 α ⁺; $\gamma\delta$ is CD4⁻CD8 α ⁻TCR- $\gamma\delta$ ⁺.

mice expressing different levels of pT α on wild-type and *Ptcr*^{-/-} backgrounds¹⁷. Such mice confirmed an inverse correlation between double-positive cells and the representation of Foxp3⁺ cells among CD4⁺ cells: as *Ptcr*^{-/-} mice were provided with transgenic pT α , more double-positive cells developed and the regulatory:effector balance declined to normal (2.5–4.5%) (Fig. 3d). This inverse relationship was almost superimposable upon that between double-positive cells and $\gamma\delta$ cells, expressed as a percentage of total thymocytes¹⁷ (Fig. 3e), but differed completely from the relationship of double-positive cells to unfractionated CD4⁺ thymocytes (Fig. 3f).

pT α expression might dictate T_R-cell representation via a cell-autonomous effect or an effect *in trans*, as exemplified by *trans*-conditioning⁷. To distinguish between these, *Ptcr*^{-/-} thymocyte differentiation was examined in sub-lethally irradiated mice reconstituted with admixed *Ptcr*^{-/-} and wild-type bone marrow, using congenic markers for each donor and the host. After 10–14 weeks, signatory *Ptcr*^{-/-} thymocyte development was evident in a large double-negative compartment, a relatively small double-positive compartment, and some single-positive cells, but this was set against a backdrop of a large double-positive pool supplied by the progeny of wild-type donor and/or recovering host bone marrow (Fig. 4a). In this context, the regulatory:effector ratio among *Ptcr*^{-/-} thymocytes was quantified as the percentage of CD4⁺ cells displaying the T_R phenotype. At minimum there was partial rescue, and three of seven mice displayed equivalent regulatory:effector ratios (for example, 3.9% and 4.5%, respectively) in cells derived from *Ptcr*^{-/-} and wild-type progenitors. Thus, the establishment of a normal ratio can be rescued *in trans* without any obligatory cell-autonomous

provision of pT α . Because normal double-positive differentiation is associated with increased thymus size, one might consider size as the primary influence on T_R-cell differentiation. Indeed, a thymus lacking a large double-positive pool does not properly support thymocyte proliferation¹⁸. However, reflecting variable susceptibility to irradiation, reconstituted thymi varied in size from ~200% to <10% of normal, yet the regulatory:effector cell ratio was similar wherever double-positive cells composed the major thymocyte subset (Fig. 4b). Likewise, $\gamma\delta$ cells differentiating from *Ptcr*^{-/-} progenitors in the context of large double-positive compartments regained the aggregate effector gene profile of wild-type $\gamma\delta$ cells (data not shown).

Of the four early thymocyte double-negative (DN) stages, *trans*-conditioning targets DN1/DN2 cells (ref. 7 and data not shown) before TCR gene rearrangement. Reflecting this, full genome profiling revealed 360 RNAs ≥ 2 -fold upregulated and 97 RNAs reciprocally downregulated in DN2 cells from wild-type compared with *Ptcr*^{-/-} mice (data not shown). Upregulated RNAs included the *ICER* (also known as *Crem*) transcriptional repressor and many other genes (for example, *Nor1*, *Ccl4*, *Rgs1*, *Nurr1* and *Bcl2a1*) that originally identified *trans*-conditioning⁷ (Supplementary Table 3). To further characterize *trans*-conditioned cells, we exploited a transgenic mouse in which the *ICER* promoter drives *lacZ* expression⁷. Numerous *trans*-conditioned genes were expressed more strongly by flow-sorted ICER^{hi} DN2 cells, which also appeared as large blasts compared with smaller ICER^{lo} cells (Supplementary Fig. 3). This suggested that size could separate conditioned (large) DN2-L thymocytes from unconditioned (small) DN2-S thymocytes using

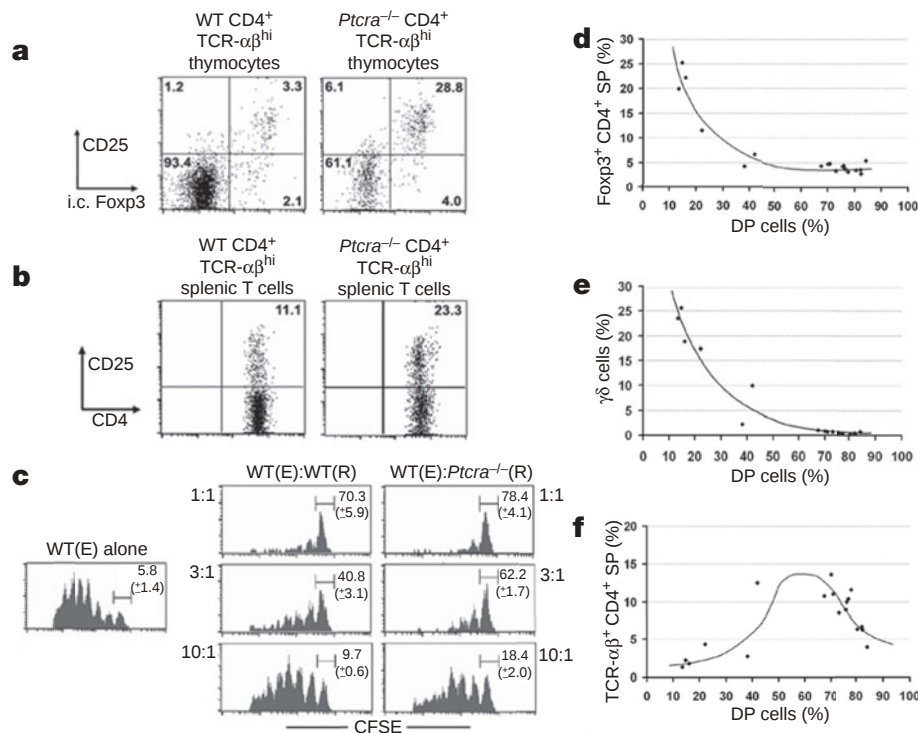


Figure 3 | *Ptcr*^{-/-} CD4⁺TCR- $\alpha\beta$ ^{hi} thymocytes and splenocytes are enriched for Foxp3⁺ regulatory cells. **a**, **b**, Representative ($n > 6$) CD25 or intracellular (i.c.) Foxp3 plots of wild-type or *Ptcr*^{-/-} CD4⁺TCR- $\alpha\beta$ ^{hi} single-positive thymocytes or splenocytes. Quadrant percentages are indicated. **c**, CFSE-labelled wild-type CD4⁺CD25⁻ splenic effector T cells (5×10^4) (WT(E)) were cultured (in triplicate) with either wild-type (WT(R)) or *Ptcr*^{-/-} (*Ptcr*^{-/-}(R)) splenic CD4⁺CD25⁺ cells in the presence of soluble anti-CD3 ϵ monoclonal antibody (+antigen presenting

cells) for 72 h. Average percentage of undivided cells ($\pm$ s.d.) is indicated. **d–f**, Adult mice thymus ($n = 18$) from *Ptcr*^{-/-}.*Lck-Ptcr-transgene*^{+/-} $\times$ *Ptcr*^{-/-} crosses were analysed for percentage CD4⁺CD8 $\alpha\beta$ ⁺ cells (DP cells (%)), percentage TCR- $\gamma\delta$ ⁺ cells ($\gamma\delta$ cells (%)) and percentage CD4⁺CD8 $\alpha\beta$ ⁻ TCR- $\alpha\beta$ ⁺ single-positive cells (TCR- $\alpha\beta$ ⁺CD4⁺ SP (%)). The percentage of TCR- $\alpha\beta$ ⁺CD4⁺ single-positive cells that was also Foxp3⁺ was also assessed (Foxp3⁺CD4⁺ SP (%)). Lines in the scatter graphs are suggestive of possible correlation.

unmanipulated wild-type mice. Although size-separation is inevitably imprecise (for example, small cells will include conditioned cells that recently underwent cytokinesis), profiling showed that the aggregate transcriptomes of DN2-L and DN2-S cells were distinct, and by hierarchical clustering analysis DN2-S cells resembled *Ptcr*^{-/-} DN2 cells more than DN2-L cells (Supplementary Fig. 4), consistent with reduced *trans*-conditioning.

To monitor the developmental potentials of the purified DN2 subsets, paired fetal thymic organ cultures (FTOCs) were used. DN2-S cells differentiated slowly, and by day 8 the double-positive pool was much smaller than in the faster differentiating DN2-L cultures (Fig. 4c). However, SP4 cells arose, and using three independent paired FTOCs harvested on days 13–21, we compared the expression of *Cd25*, *Foxp3* and seven genes (*Gitr*, *Lag3*, *Ppe*, *Ctla4*, *Cd27*, *Cd28* and *Cd122*) reportedly enriched in wild-type CD25^{hi} versus CD25^{lo} SP4 cells^{19–22}. All were overexpressed in CD25^{hi} SP4 cells from wild-type and *Ptcr*^{-/-} mice, and in SP4 cells derived from DN2-S relative to DN2-L cultures, even without prior CD25 subsetting. Differential expression of *Foxp3* was highly significant at $P < 0.01$, as measured by two-tailed *t*-test. Conversely, cyclin D3 was enriched in wild-type CD25^{lo} SP4 cells and in DN2-L-derived SP4 cells (Fig. 4d, e). In sum, from wild-type DN2 thymocytes (that temporally precede pre-TCR selection into the double-positive pool and subsequent TCR-mediated selection) one can purify DN2-S cells that show reduced evidence of *trans*-conditioning and that differentiate into

cells enriched in hallmarks of the T_R-cell phenotype. Unlike other models, these data readily explain the bias towards regulatory differentiation in the $\alpha\beta$ and $\gamma\delta$ T-cell compartments of mutant mice lacking normal double-positive compartments. A revised model that may be useful as a basis for further experimentation (Fig. 4f; see also Supplementary Fig. 5) proposes that *trans*-conditioning of DN2 progenitors promotes replication and blasting (DN2-L cells) before further pre-TCR-dependent proliferation at the DN4 to double-positive stage. Their progeny that engage partial agonists positively select as CD4⁺ effectors, whereas those engaging agonists on epithelial or bone-marrow-derived cells die by negative selection. Conversely, some DN2 cells will ordinarily receive less *trans*-conditioning, perhaps stochastically, or because of anatomical (niches) and/or cellular heterogeneity. Such cells differentiate as DN2-S cells without overt expansion before pre-TCR-dependent proliferation. The distinct gene expression of DN2-S-derived double-positive cells then dictates that those engaging agonists, for example on epithelial cells, select as T_R cells, rather than deleting. Thus, whereas effector cell differentiation in *Ptcr*^{-/-} mice is impaired by reduced cell-autonomous expansion and *trans*-conditioning, only the former impairs regulatory cell differentiation. Mathematical models are consistent with unconditioned cells being a major source of T_R cells for which agonist selection^{4–6} would be a final, essential manifesting step. It has also been noted that the promotion of T_R-cell differentiation by SHP-1 deficiency is consistent with T_R cells arising from a distinct progenitor

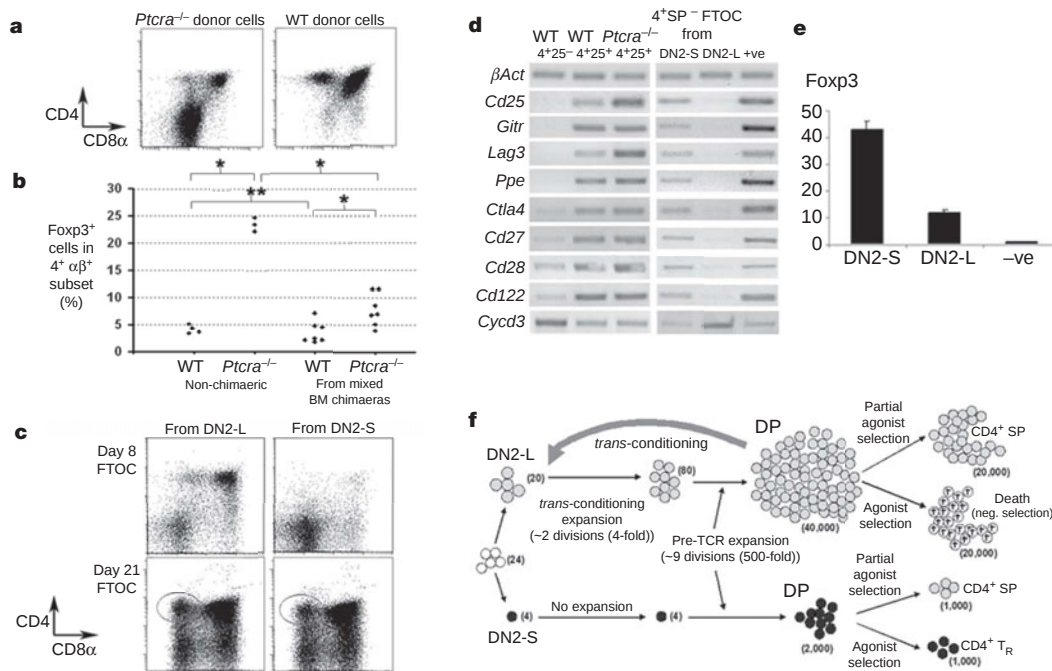


Figure 4 | *Trans*-conditioning regulates the differential capacity of early thymic progenitors to develop into T cells with regulatory characteristics. **a**, Representative ($n = 7$) CD4/CD8 α plots of wild-type (Ly5.1) and *Ptcr*^{-/-} (Ly5.2) donor thymocytes from a 12-week bone-marrow chimaera. **b**, Graph displaying percentage of CD4⁺ TCR- $\alpha\beta$ ⁺ thymocytes that are Foxp3⁺ from wild-type and *Ptcr*^{-/-} non-chimaeric mice, and from mixed wild-type/*Ptcr*^{-/-} bone-marrow chimaeras described in **a**. *P*-values are shown; asterisk, $P < 0.05$; double asterisk, not significant at $P = 0.05$. **c**, CD4/CD8 α plots of wild-type DN2 thymocytes (CD4⁺CD8 α ⁺TCR- $\alpha\beta$ ⁺TCR- $\gamma\delta$ ⁻B220⁻NK1.1⁻DX5⁻CD25⁺CD44⁺) size-sorted into the largest 20%, DN2-L, and the smallest 20%, DN2-S, and cultured as 8- or 21-day FTOC (1×10^4 cells per lobe). CD4⁺ single-positive (SP) gates for analyses in panels **d** and **e** are shown (ovals). **d**, Representative ($n = 3$) RT-PCR for regulatory T-cell-associated genes in: wild-type CD4⁺CD8 α ⁺CD25⁻ cells (WT 4⁺25⁻); wild-type CD4⁺CD8 α ⁺CD25⁺ cells (WT 4⁺25⁺); *Ptcr*^{-/-} CD4⁺CD8 α ⁺CD25⁻ cells (*Ptcr*^{-/-} 4⁺25⁻); and CD4⁺CD8 α ⁺ cells from

13-day FTOC (as described for **c**; DN2-S and DN2-L). **e**, Real-time PCR for Foxp3 in CD4⁺CD8 α ⁺ cells from 21-day FTOC (as described for **c**). Units for the y axis are Foxp3/HPRT ($\times 10^4$). cDNA from *Rag1*^{-/-} thymus is a negative control (-ve). Results are significantly different by both two-tailed Student's *t*-test (27.1, $P < 0.01$; SPSS v.13.0 software) and non-parametric Mann-Whitney *U*-test ($P < 0.05$). Data show mean $\pm$ s.d. **f**, Revised working model for thymic T_R-cell development. DN2-L cells upregulate gene expression and undergo expansion by *trans*-conditioning (mediated by WT DP cells), whereas DN2-S cells do neither. DN2-L cells develop into conventional double-positive cells that either positively select on partial agonist or negatively select on full agonist. In contrast, DN2-S cells produce a minor, yet distinct double-positive subset that differentiates to CD4⁺ T_R cells when encountering full agonist on epithelial cells. This scheme predicts that pT α -deficiency will reduce total thymocyte output and alter the effector:regulatory T-cell balance towards T_R cells. Numbers indicate relative thymocyte number (see Supplementary Fig. 5).

pool²³, which might additionally explain the seeming ceiling on T_R-cell numbers even in TCR-transgenic mice where there is increased opportunity for agonist selection^{24,25}.

That the regulatory:effector balance increases whenever thymic double-positive pools are reduced may partially explain T_R-cell enhancement in pregnancy, because double-positive cells are depleted, purportedly by steroid hormones²⁶. Chronic infections may also deplete thymic double-positive cells, and it will be intriguing if there are parallel increases in regulatory potential, possibly to limit immunopathology. Although lymphotoxin determines several aspects of *trans*-conditioning of $\gamma\delta$ cells⁷, its contribution to the regulatory:effector balance is less clear. Although its deficiency may reduce conditioning, thus promoting T_R-cell differentiation, its impairment of AIRE expression may limit T_R-cell selection. Thus, future studies must aim to define the molecules that act in early thymocyte differentiation to determine the regulatory:effector balance in the $\alpha\beta$ and $\gamma\delta$ T-cell compartments.

Note added in proof: Since the preparation of this paper, the overrepresentation of T_R cells in *Ptcr*^{-/-} mice has also been reported, with additional explanations³⁰.

METHODS

Mice. Mice have been described previously^{16,17,27–29}. For generation of mixed bone-marrow chimaeras, 5×10^6 bone-marrow cells from wild-type (C57/BL6 Ly5.1⁺Thy1.2⁺) or *Ptcr*^{-/-} (Ly5.2⁺Thy1.2⁺) mice were mixed at wild-type:*Ptcr*^{-/-} ratios of 1:1, 1:5 or 1:10 and injected intravenously into irradiated (900 rad) wild-type recipient mice (C57BL/6 Ly5.2⁺Thy1.1⁺). Analysis was at 12 weeks.

Flow cytometry. Antibodies were from eBioscience or BD Pharmingen. Pre-enriched double-negative thymocytes were obtained by complement lysis (low-tox-M rabbit complement, Cedarlane) of CD4⁺ and CD8⁺ cells using hybridomas; RL172; 31M (gift of R. Ceredig). Intracellular staining of Foxp3 used the eBioscience protocol (number 72-5775-40). Visualization of β -galactosidase expression has been described⁷.

RT-PCR. DNase-treated (RQ1, Promega) total RNA became oligodT-primed cDNA using the PowerScript cDNA synthesis kit (Clontech). Semi-quantitative PCR was performed on cDNA equivalent to $1\text{--}2 \times 10^3$ cells (Advantage2, Clontech) on a Tetrad DNA engine (MJ Research). Real-time PCR was performed on an ABI Prism 7700 Sequence Detection System (PE Applied Biosystems) using SYBR GreenER qPCR SuperMix (Applied Biosystems). For primers see Supplementary Table 4.

In vitro regulatory T-cell assay. Sorted, CFSE-labelled (Molecular Probes), wild-type CD4⁺CD25⁻ T cells (5×10^4 per well) from lymph node/spleen were targets. Sorted, DDAO-SE-labelled (Molecular Probes), wild-type or *Ptcr*^{-/-} lymph node/splenic CD4⁺CD25⁺ or $\gamma\delta$ cells were potential regulators. Red-blood-cell-depleted mitomycin-C-treated (Sigma) splenocytes were used as antigen presenting cells (5×10^4 per well). Cells were mixed in 96-well round-bottom plates, activated with soluble anti-CD3 ϵ monoclonal antibody (2C-11, Pharmingen) for 72–84 h, and analysed by flow cytometry.

Fetal thymic organ cultures. C57BL/6 day-15 fetal thymic lobes were treated for 5 days with 2-deoxyguanosine (1.35 mM) in FTOC media (RPMI 10% FCS, 50 μ M β -mercaptoethanol, 10 mM HEPES, L-Glut/Pen/Strep). Hanging-drop cultures in Terasaki plates with 1×10^4 thymocytes for 24 h were followed by 8–21-day culture on nucleopore filters in FTOC media, before FACS analysis.

Transcription profiling. Total RNA was amplified and labelled by MessageAmp II aRNA amplification kit and Illumina TotalPrep RNA amplification kit (Ambion). cRNA was hybridized to Sentrix Mouse-6 Expression Beadchips (Illumina). Analysis used Genespring GX7.3 (Agilent Technologies).

Received 6 September; accepted 24 October 2006.

Published online 12 November 2006.

- Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003).
- Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4⁺CD25⁺ regulatory T cells. *Nature Immunol.* **4**, 330–336 (2003).
- Khattari, R., Cox, T., Yasayko, S. A. & Ramsdell, F. An essential role for Scurfin in CD4⁺CD25⁺ T regulatory cells. *Nature Immunol.* **4**, 337–342 (2003).
- Jordan, M. S. *et al.* Thymic selection of CD4⁺CD25⁺ regulatory T cells induced by an agonist self-peptide. *Nature Immunol.* **2**, 301–306 (2001).
- Apostolou, I., Sarukhan, A., Klein, L. & von Boehmer, H. Origin of regulatory T cells with known specificity for antigen. *Nature Immunol.* **3**, 756–763 (2002).

- Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
- Silva-Santos, B., Pennington, D. J. & Hayday, A. C. Lymphotoxin-mediated regulation of $\gamma\delta$ cell differentiation by $\alpha\beta$ T cell progenitors. *Science* **307**, 925–928 (2005).
- Havran, W. L. & Allison, J. P. Origin of Thy-1⁺ dendritic epidermal cells of adult mice from fetal thymic precursors. *Nature* **344**, 68–70 (1990).
- Girardi, M. *et al.* Resident skin-specific $\gamma\delta$ T cells provide local, nonredundant regulation of cutaneous inflammation. *J. Exp. Med.* **195**, 855–867 (2002).
- Wise, E., Lewis, J., LeClercq, G., Tigelaar, R. E. & Hayday, A. C. Molecular analysis of skin-resident T cells clarifies functional potential, heterogeneity, strain variation and major differences to gut intraepithelial T cells. *J. Immunol.* (submitted).
- Suri-Payer, E., Amar, A. Z., Thornton, A. M. & Shevach, E. M. CD4⁺CD25⁺ T cells inhibit both the induction and effector function of autoreactive T cells and represent a unique lineage of immunoregulatory cells. *J. Immunol.* **160**, 1212–1218 (1998).
- Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor α -chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol.* **155**, 1151–1164 (1995).
- Livak, F., Wilson, A., MacDonald, H. R. & Schatz, D. G. $\alpha\beta$ lineage-committed thymocytes can be rescued by the $\gamma\delta$ T cell receptor (TCR) in the absence of TCR β chain. *Eur. J. Immunol.* **27**, 2948–2958 (1997).
- Passoni, L. *et al.* Intrathymic δ selection events in $\gamma\delta$ cell development. *Immunity* **7**, 83–95 (1997).
- Buer, J., Aifantis, I., DiSanto, J. P., Fehling, H. J. & von Boehmer, H. Role of different T cell receptors in the development of pre-T cells. *J. Exp. Med.* **185**, 1541–1547 (1997).
- Fehling, H. J., Krotkova, A., Saint-Ruf, C. & von Boehmer, H. Crucial role of the pre-T-cell receptor α gene in development of $\alpha\beta$ but not $\gamma\delta$ T cells. *Nature* **375**, 795–798 (1995).
- Gibbons, D. *et al.* The biological activity of natural and mutant pT α alleles. *J. Exp. Med.* **194**, 695–703 (2001).
- Prockop, S. E. & Petrie, H. T. Regulation of thymus size by competition for stromal niches among early T cell progenitors. *J. Immunol.* **173**, 1604–1611 (2004).
- Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. & Rudensky, A. Y. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nature Immunol.* **6**, 1142–1151 (2005).
- Zelenika, D. *et al.* Regulatory T cells overexpress a subset of Th2 gene transcripts. *J. Immunol.* **168**, 1069–1079 (2002).
- Read, S., Malmstrom, V. & Powrie, F. Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25⁺CD4⁺ regulatory cells that control intestinal inflammation. *J. Exp. Med.* **192**, 295–302 (2000).
- Takahashi, T. *et al.* Immunologic self-tolerance maintained by CD25⁺CD4⁺ regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J. Exp. Med.* **192**, 303–310 (2000).
- Carter, J. D., Calabrese, G. M., Naganuma, M. & Lorenz, U. Deficiency of the Src homology region 2 domain-containing phosphatase 1 (SHP-1) causes enrichment of CD4⁺CD25⁺ regulatory T cells. *J. Immunol.* **174**, 6627–6638 (2005).
- van Santen, H. M., Benoist, C. & Mathis, D. Number of T reg cells that differentiate does not increase upon encounter of agonist ligand on thymic epithelial cells. *J. Exp. Med.* **200**, 1221–1230 (2004).
- Lerman, M. A., Larkin, J. III, Cozzo, C., Jordan, M. S. & Caton, A. J. CD4⁺CD25⁺ regulatory T cell repertoire formation in response to varying expression of a neo-self-antigen. *J. Immunol.* **173**, 236–244 (2004).
- Aluvihare, V. R., Kallikourdis, M. & Betz, A. G. Regulatory T cells mediate maternal tolerance to the fetus. *Nature Immunol.* **5**, 266–271 (2004).
- Futterer, A., Mink, K., Luz, A., Kosco-Vilbois, M. H. & Pfeffer, K. The lymphotoxin β receptor controls organogenesis and affinity maturation in peripheral lymphoid tissues. *Immunity* **9**, 59–70 (1998).
- Philpott, K. L. *et al.* Lymphoid development in mice congenitally lacking T cell receptor $\alpha\beta$ -expressing cells. *Science* **256**, 1448–1452 (1992).
- Mombaerts, P. *et al.* Mutations in T-cell antigen receptor genes α and β block thymocyte development at different stages. *Nature* **360**, 225–231 (1992).
- Bosco, N., Agenes, F., Rolink, A. G. & Ceredig, R. Peripheral T cell lymphopenia and concomitant enrichment in naturally arising regulatory T cells: The case of the pre-T α gene-deleted mouse. *J. Immunol.* **177**, 5014–5023 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank the Wellcome Trust for support, and W. Turnbull, E. Wise, G. Warnes, T. Tree, L. Graca, A. Aguirre, P. Pereira, D. Gibbons, S. Haque, Jian-Guo Chai, H. Brady, C. Langford, P. Ellis, J. Lewis, M. Girardi and R. Tigelaar for assistance, support and discussions. T.S. is supported by a Ph.D. scholarship from the Boehringer Ingelheim Fonds.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J.P. (d.pennington@qmul.ac.uk) or A.C.H. (adrian.hayday@kcl.ac.uk).

LETTERS

The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA

Qiaozhen Ye¹, Robert M. Krug² & Yizhi Jane Tao¹

Influenza A viruses pose a serious threat to world public health, particularly the currently circulating avian H5N1 viruses. The influenza viral nucleoprotein forms the protein scaffold of the helical genomic ribonucleoprotein complexes, and has a critical role in viral RNA replication¹. Here we report a 3.2 Å crystal structure of this nucleoprotein, the overall shape of which resembles a crescent with a head and a body domain, with a protein fold different compared with that of the rhabdovirus nucleoprotein^{2,3}. Oligomerization of the influenza virus nucleoprotein is mediated by a flexible tail loop that is inserted inside a neighbouring molecule. This flexibility in the tail loop enables the nucleoprotein to form loose polymers as well as rigid helices, both of which are important for nucleoprotein functions. Single residue mutations in the tail loop result in the complete loss of nucleoprotein oligomerization. An RNA-binding groove, which is found between the head and body domains at the exterior of the nucleoprotein oligomer, is lined with highly conserved basic residues widely distributed in the primary sequence. The nucleoprotein structure shows that only one of two proposed nuclear localization signals are accessible, and suggests that the body domain of nucleoprotein contains the binding site for the viral polymerase. Our results identify the tail loop binding pocket as a potential target for anti-viral development.

We expressed the nucleoproteins of several different influenza A virus strains in *Escherichia coli* and insect cells. These recombinant nucleoproteins form non-homogeneous assemblies ranging from monomers to large insoluble aggregates. On the basis of results from gel filtration chromatography, the most abundant molecular species are small oligomers of two to three hundred kilodaltons (Fig. 1). The degree of nucleoprotein oligomerization also varies between influenza A virus strains (Fig. 1). When the nucleoproteins collected from the monomeric peak position were reapplied to the same gel filtration column, most of the nucleoprotein re-appeared as oligomers. Therefore, monomeric and oligomeric forms are interchangeable, and different nucleoprotein assembly states are in dynamic equilibrium. Electron microscopy of the influenza A/WSN/33 virus nucleoprotein taken from the oligomeric peak fraction of the gel filtration (~150 kDa) shows many closed ring structures (Fig. 2a). Although trimers are most common, larger oligomers and monomers are also observed. When viewed along the symmetry axis of an oligomer, each nucleoprotein molecule has the shape of an ellipsoid, with a small lateral surface area buried between neighbouring nucleoprotein subunits.

Although we were unable to isolate nucleoprotein in a single molecular form, nucleoprotein samples collected from the oligomeric peak fractions of a gel filtration column produced thin-plate crystals that diffract to 3.0 Å at synchrotron radiation sources. These thin-plate crystals belong to the space group *C*222₁ with one complete

trimer per asymmetric unit. The three-fold non-crystallographic symmetry (NCS) does not fall on any of the major crystallographic symmetry axes. Using multiple isomorphous replacement and anomalous scattering (MIRAS), we determined the crystal structure of influenza A virus nucleoprotein (strain A/WSN/33) to 3.2 Å resolution (Supplementary Table 1 and Supplementary Fig. 1). Of the three subunits in an NCS trimer, subunits A and B each contain 426 amino acid residues in the final model, including residues 21–72, 92–202, 213–396, 402–428 and 438–489. Subunit C contains ten more residues, as residues 203–212 are ordered in this subunit. There are six cysteine residues in each nucleoprotein monomer, but none is close enough for disulphide bond formation. The structure of nucleoprotein is mostly α -helical and each monomer has a curved, crescent-like shape (Fig. 3; see also Supplementary Fig. 2). The overall structure can be divided into two domains: a head and a body, which are joined together by the polypeptide chain at three regions. The head and body domains are formed by non-contiguous polypeptide regions: the head domain is formed by residues 150–272 and 438–452, and the body domain consists of the three polypeptide segments 21–149, 273–396 and 453–489. Consequently, the overall fold and topology of the influenza virus nucleoprotein are different from that of the rhabdovirus nucleoproteins in which the two domains are formed by collinear polypeptide sequences^{2,3}.

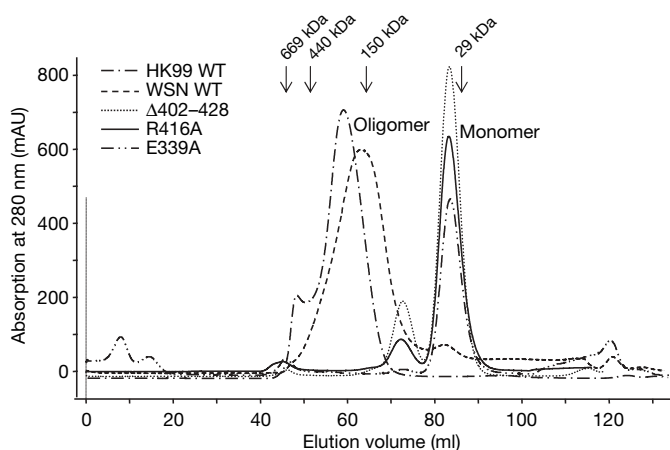


Figure 1 | Nucleoprotein oligomerization behaviour shown by chromatogram. Wild-type nucleoprotein from A/Hong Kong/1074/99 (H9N2) (HK99 WT), wild-type nucleoprotein from A/WSN/33 (WSN WT), and Δ 402–428, R416A and E339A mutant nucleoprotein from A/WSN/33 (H1N1) were eluted from a Superdex 200 size-exclusion column. The peak positions of protein standards are marked by arrows. Nucleoprotein monomer and oligomer peaks are also marked.

¹Department of Biochemistry and Cell Biology, Rice University, 6100 Main Street, MS140, Houston, Texas 77005, USA. ²Institute for Cellular and Molecular Biology, University of Texas at Austin, 2500 Speedway, Austin, Texas 78712-1095, USA.

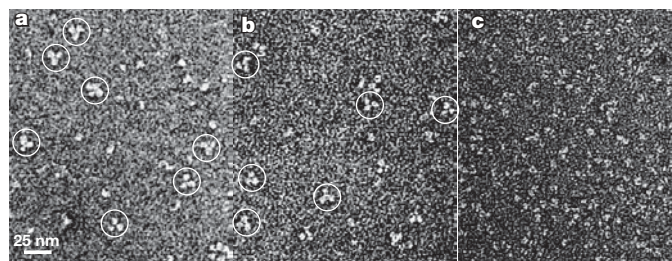


Figure 2 | Electron microscopy images of influenza virus nucleoprotein (strain A/WSN/33). **a**, Recombinant nucleoprotein. Nucleoprotein trimers are highlighted by white circles. A small number of nucleoprotein monomers and larger oligomers are also observed. **b**, Nucleoprotein in a complex with a

24-nucleotide ssRNA. The nucleoprotein:RNA complex was prepared by mixing nucleoprotein and RNA at the $\sim 1:2$ ratio. **c**, The R416A mutant nucleoprotein. All samples were stained using uranyl formate as described in Methods. Scale bar, 25 nm.

Each nucleoprotein contains a tail loop, formed by residues 402–428, at the back of the crescent-shaped molecule (Fig. 3d). The polypeptide connecting the tail loop to the main body of the nucleoprotein structure is disordered. The tail loop has a rather extended conformation and interacts with a neighbouring molecule in an NCS trimer. This tail loop is inserted into the body domain of a neighbouring molecule in an anticlockwise direction when viewed along the three-fold axis from the side of the head domain. Intra-molecular interactions for the tail loop are not possible because the distance is too far for the tail loop to reach the binding pocket located on the opposite side of the same molecule. The tail loop contains three β -strands, with the last two forming an intra-molecular β -hairpin. The tail loop from one molecule makes extensive interactions with a neighbouring subunit through several types of molecular forces, including intermolecular β -sheets (for example, $\beta 7$ with $\beta 8$; $\beta 6$ with $\beta 4$), hydrophobic interactions (for example, Pro 453(A), Phe 420(B) and Ile 453(A)), and salt bridges (for example, Glu 339(A) with Arg 416(B); Arg 422(B) with Glu 449(A)) (Fig. 4). (The last capitalized letter for amino acids indicates the molecular origin of that

residue; for example, (A) indicates subunit A.) The interactions mediated by the tail loop are almost equivalent among the three NCS-related subunits (Supplementary Table 2). In addition, because the oligomerization tail loop is loosely attached to the rest of the nucleoprotein through highly flexible linkers that are disordered in electron density maps, the tail loop can make similar inter-subunit interactions in both closed rings (as observed by electron microscopy in Fig. 2a) and helical ribonucleoprotein (RNP)-like structures, as has been proposed for rhabdovirus nucleocapsid proteins^{2,3}. The carboxy-terminal region of each monomer (residues 440–end) also makes some inter-subunit contacts, but these are far fewer in number compared to the tail loop (Supplementary Table 2). In addition, many such interactions among the three subunits are not conserved and therefore also suggest that the molecular interface mediated by the C-terminal region is labile and not critical for nucleoprotein oligomerization.

To determine whether the tail loop is essential for nucleoprotein oligomerization, we made two single residue mutants, R416A and E339A, and a deletion mutant, $\Delta 402$ –428, that removes the tail loop.

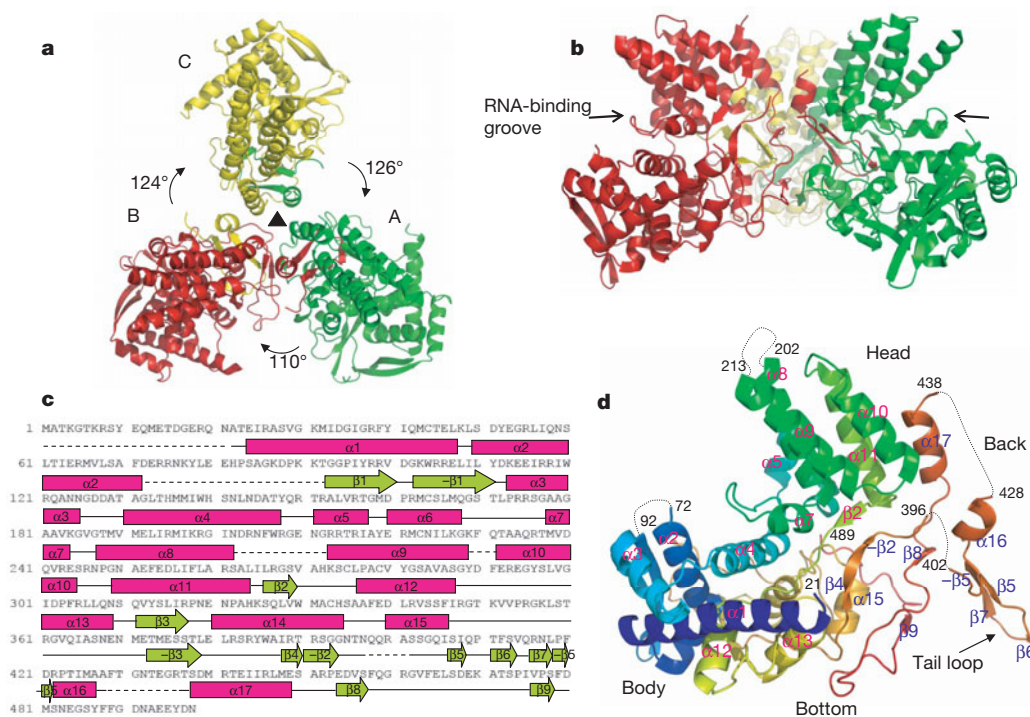


Figure 3 | Nucleoprotein crystal structure. **a**, Nucleoprotein trimer viewed along the NCS three-fold axis, with three subunits shown in different colours. The rotation angles that relate the three subunits are marked. **b**, Side view of a nucleoprotein trimer, with the three subunits coloured as in **a**. **c**, Nucleoprotein secondary structure assignment (subunit B). α -Helices

and β -strands are shown by rods and arrows, respectively. β -Strands with opposite signs form antiparallel β -strands. **d**, Nucleoprotein subunit B. The polypeptide is coloured continuously from blue to red. Residues preceding and succeeding disordered regions are marked and connected by dotted lines.

Residues R416 and E339 form an inter-subunit salt bridge. When the oligomerization behaviour of R416A, E339A and $\Delta 402$ –428 was evaluated by gel filtration chromatography, all three mutants did not self-associate and formed only monomers (Fig. 1). Electron microscopy of the R416A mutant protein confirmed that only monomers were formed (Fig. 2c), in contrast to wild-type nucleoprotein where trimers and larger oligomers were frequently observed (Fig. 2a). Proper self-association is important for the biological activities of nucleoprotein, because the R416A mutant was shown to be defective in RNA binding and also in viral RNA synthesis⁴. This observation may indicate that the displacement of the tail loop from its binding pocket causes significant structural rearrangements in nucleoprotein. We anticipate that chemical compounds which competitively displace the tail loop from its binding pocket would interfere with viral genome replication, and therefore serve as promising leads for anti-influenza drug development.

The nucleoprotein crystal structure shows that a deep groove between the head and body domain, which is at the exterior of nucleoprotein oligomers, is probably the RNA-binding site (Figs 3b, d and 5). The surface of the groove is made up of a large number of basic residues, including R65, R150, R152, R156, R174, R175, R195, R199, R213, R214, R221, R236, R355, K357, R361 and R391, that are likely to function in RNA binding by interacting with the phosphodiester backbone. An aromatic residue, Y148, is found at one end of the groove and may stack with a nucleotide base (Fig. 5b). Eleven out of the sixteen basic residues observed at the putative RNA-binding surface are conserved among influenza A, B and C viruses (of strains A/WSN/33, B/Ann Arbor/1/66, and C/California/78, respectively), further supporting a role in RNA binding. The aromatic residue Y148 is also conserved. Because these positively charged residues are widely distributed along the nucleoprotein polypeptide, previous mutagenesis studies have identified only a small subset of the basic amino acids in the RNA-binding groove^{4–6}. With RNA bound to the groove at the outer periphery of oligomeric nucleoprotein, the distance between two neighbouring binding sites in a nucleoprotein trimer is approximately 70 Å, consistent with the stoichiometry of one nucleoprotein per 24 nucleotides of RNA, as previously determined⁷.

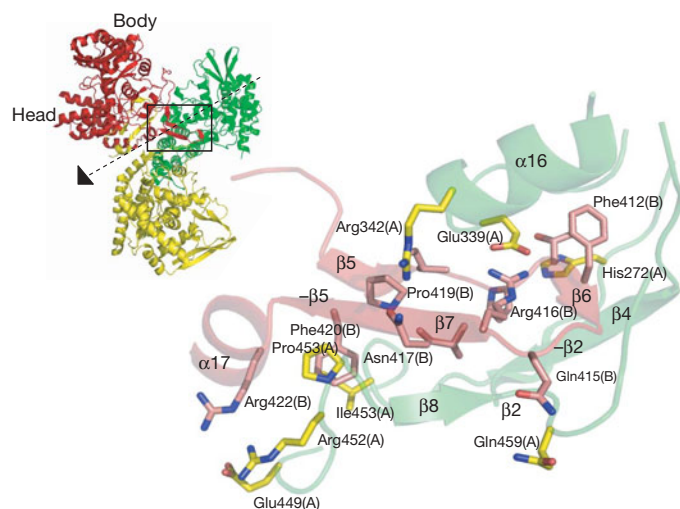


Figure 4 | Inter-subunit interactions mediated by the tail loop. Subunits A and B are shown in green and red, respectively. Amino acid side chains are drawn only for those residues making important interactions (for a thorough list, see Supplementary Table 2). The last capitalized letter in amino acid labels is used to show the molecular origin of that residue; for example, Glu 339(A) means residue Glu 339 from subunit A. The tail loop region is marked on a nucleoprotein trimer (boxed region), the three-fold axis of which is shown by a dotted arrow. At this orientation, the nucleoprotein trimer is tilted with the body domain situated on top.

To determine whether our nucleoprotein preparation possesses RNA-binding activity, we carried out gel-shift experiments using a 24-nucleotide single-stranded (ss)RNA, which is predicted to bind only a single nucleoprotein molecule^{7,8}. At a nucleoprotein:RNA ratio of 1:1 and 2:1, 95% and 100% of the RNA, respectively, is bound to nucleoprotein (Supplementary Fig. 3). These nucleoprotein:RNA complexes form trimeric oligomers like those observed for RNA-free nucleoprotein (Fig. 2b), indicating that RNA binding itself does not affect nucleoprotein–nucleoprotein association, and that RNA-bound nucleoprotein is likely to use a similar set of oligomerization forces as those in free nucleoprotein. The polymerization of nucleoprotein into large RNP-like structures is likely to be the direct consequence of cooperative binding of nucleoprotein to long RNA molecules. Because the oligomerization tail loop is flexibly attached to the nucleoprotein body domain, we anticipate that many of the inter-subunit interactions now observed in the nucleoprotein crystal structure would be preserved in the continuous nucleoprotein:RNA helical fibre in RNPs, and that similar nucleoprotein–nucleoprotein interactions would also be maintained when the RNP is partially unwound during RNA synthesis. On the basis of the observation that RNA-bound nucleoprotein appears to curve slightly towards the interior in a mini-RNP⁹, it is possible that RNA binding to the exterior groove in the nucleoprotein induces a small change in the relative orientation of the body and head domains. Our results indicate that RNA would be largely exposed on the external surface of nucleoprotein molecules in the RNP, consistent with the fact that influenza virus RNA coated with nucleoprotein molecules is RNase-sensitive¹⁰. In contrast, the rhabdovirus and bornavirus nucleoproteins have their RNA-binding groove facing the interior^{2,3,11}, and would be expected to protect the bound RNA from RNase digestion, as has been observed¹².

The nucleoprotein structure indicates that the non-conventional nuclear localization signal (nNLS: ³Sx**G**T**K**R**S**Yxx**M**, with important residues highlighted in bold)^{13–15} previously identified at the amino terminus of the influenza nucleoprotein can function as its NLS. Because influenza viral RNA replication and RNP assembly both take place in the nucleus¹, newly synthesized nucleoproteins have to be transported into the nucleus. The nNLS sequence is located at the outer periphery of the nucleoprotein trimer where it is highly accessible to solvent (Fig. 3d). In addition, this region is disordered in electron density maps, and this flexibility should allow the nNLS to fit easily into the substrate recognition pocket on importin- α . Another NLS, a classical bipartite NLS (cNLS: ¹⁹⁸K**R**G**I**N**D**R**N**F**W**R**G** **E**N**G**R**K**T**R**) that is located near the middle of the polypeptide, has been described¹⁶. The nucleoprotein structure suggests that this sequence does not function as a NLS, because the ¹⁹⁸K**R** and the ²¹³R**K**T**R** motifs are ~15 Å apart, a distance far shorter than the

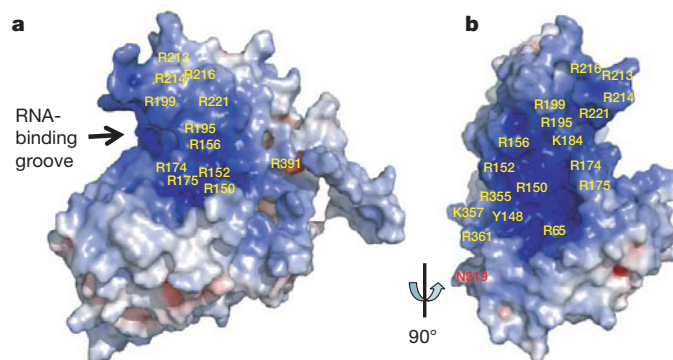


Figure 5 | RNA-binding by nucleoprotein. **a**, **b**, Electrostatic potential distribution for nucleoprotein (subunit B) identifies a potential RNA-binding site. Panel **a** is in the same orientation as the red subunit in Fig. 3b. Positive and negative electrostatic potentials are shown in blue and red, respectively.

minimal separating distance of 28 Å needed for efficient binding of a bipartite NLS¹⁷. It has also been proposed that nucleoprotein contains a cytoplasmic accumulation signal (CAS, residues 327–345), which interacts with F-actin and causes cytoplasmic retention of nucleoprotein late in infection^{16,18}. However, this region is buried internally in the nucleoprotein structure and in this conformational state would not be able to bind F-actin.

The nucleoprotein structure provides a framework for delineating the mechanism by which nucleoprotein interacts with the viral RNA polymerase to promote viral RNA replication. The interaction of the nucleoprotein with the viral polymerase is required for viral RNA replication, but not for the synthesis of viral messenger RNAs (transcription)^{19–21}. Previous experiments implicated three nucleoprotein regions—amino acids 1–160, 256–340 and 340–498—in binding to the PB1 and PB2 subunits of the viral polymerase²². Most of these amino acids (1–149, 273–340, 453–498) are localized in the body domain of the nucleoprotein (Fig. 3c, d). We propose that conserved amino acid regions in the body domain that are on the surface of nucleoprotein mediate nucleoprotein–polymerase interactions that are crucial for viral RNA replication. This hypothesis is consistent with the observation that a single amino acid change of N→K in residue 319, which is located on the surface at the right-hand side of the nucleoprotein body domain (Fig. 5b), resulted in an increase of polymerase activity²³. This nucleoprotein structure will enable investigators to make rational mutations to determine the roles of specific nucleoprotein amino acids in functional interactions with the viral RNA polymerase.

METHODS

The nucleoprotein gene from influenza virus A/WSN/33, which codes for a 498-residue protein, was cloned into vector pET28a and expressed in *E. coli* strain Rosetta 2 (DE3) Singles competent cells (Novagen). The sequence encoding the first seven amino acid residues was deleted. Expressed nucleoprotein contains a C-terminal His-tag provided by the vector. The nucleoprotein gene from influenza virus A/Hong Kong/1074/99 (H9N2), which also encodes a protein of 498 residues, was cloned into a baculovirus vector with an extra sequence 'MPRGSHHHHHHGMASMTGGQMGRLDYDDDDKDPSSR' at the N terminus. The expressed nucleoproteins were purified using Ni-NTA affinity (Qiagen), a HiTrap heparin HP column and a Superdex 200 gel filtration column (Amersham Biosciences). Purified nucleoprotein was at least 95% pure according to Coomassie SDS–polyacrylamide gel electrophoresis and had a A_{280nm}/A_{260nm} ratio of at least 1.8.

For gel-shift experiments, a 24-nucleotide ssRNA (6.5 μM) and various amounts of nucleoprotein were incubated in 15 μl binding buffer (containing 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM dithiothreitol). The RNA substrate was a 24-nucleotide ssRNA (5'-UUUGUUACACACACACGUGUG-3') obtained by *in vitro* transcription. Reaction mixtures were incubated at 37 °C for 30 min before being subjected to non-denaturing electrophoresis on a 3–9% polyacrylamide gel.

For electron microscopy, nucleoprotein samples were stained with freshly prepared 0.75% uranyl formate solution. Images were taken using a JEOL 2010 electron microscope operating at 120 kV, with a magnification of ×50,000 and a defocus of 2 to 3 μm. The sample of nucleoprotein:RNA complex was prepared by incubating nucleoprotein (650 nM) with the 24-nucleotide ssRNA at 1:2 molar ratio.

Crystals of nucleoprotein were grown using the vapour diffusion method with the drop made of equal volume of protein solution (5 mg ml⁻¹) and well solution (containing 100 mM Tris-HCl pH 7.5, 3% PEG8000, 10% glycerol, 10 mM dithiothreitol). Diffraction data were collected from single frozen crystals. A Se derivative was obtained using SeMet-substituted nucleoproteins. Other three heavy atom derivatives were prepared by soaking native crystals overnight in stabilizing solution containing 0.5 mM heavy atom compounds, followed by back-soaking for at least 2 h. Heavy atom derivative data were collected at the f'' peak wavelength determined from fluorescence scan.

Diffraction data were collected at the Brookhaven National Light Source (BNLS) and the Cornell High Energy Synchrotron Source (CHESS). All diffraction data were processed using the program HKL2000 (ref. 24).

The nucleoprotein structure was determined by multiple isomorphous replacement with anomalous scattering (MIRAS) (Supplementary Table 1). Heavy atom sites were initially determined from the anomalous signal using the program SOLVE²⁵ and refined using SHARP²⁶. Experimental phases were

further improved by solvent flipping, histogram matching and three-fold NCS averaging. The NCS symmetry operator was determined manually and refined using the program DM²⁷. The protein model, which was built using the program O²⁸, was refined using CNS²⁹ with NCS restraint and later using REFMAC³⁰ without restraints. The polypeptide tracing was confirmed using heavy atom sites. For instance, 18 of the 19 modelled methionines coincide with strong peaks ($\sigma > 2.5$) in the anomalous difference Fourier map of the SeMet derivative. All figures were prepared using the program PYMOL (W. L. Delane, <http://www.py-mol.org>) unless noted otherwise.

Received 3 October; accepted 26 October 2006.

Published online 6 December 2006.

- Lamb, R. A. & Krug, R. M. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 1487–1531 (Lippincott-Raven, Philadelphia, 2001).
- Albertini, A. A. *et al.* Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**, 360–363 (2006).
- Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357–360 (2006).
- Elton, D., Medcalf, L., Bishop, K., Harrison, D. & Digard, P. Identification of amino acid residues of influenza virus nucleoprotein essential for RNA binding. *J. Virol.* **73**, 7357–7367 (1999).
- Albo, C., Valencia, A. & Portela, A. Identification of an RNA binding region within the N-terminal third of the influenza A virus nucleoprotein. *J. Virol.* **69**, 3799–3806 (1995).
- Kobayashi, M., Toyoda, T., Adyshev, D. M., Azuma, Y. & Ishihama, A. Molecular dissection of influenza virus nucleoprotein: deletion mapping of the RNA binding domain. *J. Virol.* **68**, 8433–8436 (1994).
- Ortega, J. *et al.* Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. *J. Virol.* **74**, 156–163 (2000).
- Compans, R. W., Content, J. & Duesberg, P. H. Structure of the ribonucleoprotein of influenza virus. *J. Virol.* **10**, 795–800 (1972).
- Area, E. *et al.* 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc. Natl Acad. Sci. USA* **101**, 308–313 (2004).
- Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* **13**, 3158–3165 (1994).
- Rudolph, M. G. *et al.* Crystal structure of the borna disease virus nucleoprotein. *Structure* **11**, 1219–1226 (2003).
- Iseni, F., Barge, A., Baudin, F., Blondel, D. & Ruigrok, R. W. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J. Gen. Virol.* **79**, 2909–2919 (1998).
- Wang, P., Palese, P. & O'Neill, R. E. The NP1/–NP1-3 (karyopherin α) binding site on the influenza A virus nucleoprotein NP is a nonconventional nuclear localization signal. *J. Virol.* **71**, 1850–1856 (1997).
- Neumann, G., Castrucci, M. R. & Kawaoka, Y. Nuclear import and export of influenza virus nucleoprotein. *J. Virol.* **71**, 9690–9700 (1997).
- Cros, J. F., Garcia-Sastre, A. & Palese, P. An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. *Traffic* **6**, 205–213 (2005).
- Weber, F., Kochs, G., Gruber, S. & Haller, O. A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology* **250**, 9–18 (1998).
- Fontes, M. R., Teh, T. & Kobe, B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin- α . *J. Mol. Biol.* **297**, 1183–1194 (2000).
- Digard, P. *et al.* Modulation of nuclear localization of the influenza virus nucleoprotein through interaction with actin filaments. *J. Virol.* **73**, 2222–2231 (1999).
- Medcalf, L., Poole, E., Elton, D. & Digard, P. Temperature-sensitive lesions in two influenza A viruses defective for replicative transcription disrupt RNA binding by the nucleoprotein. *J. Virol.* **73**, 7349–7356 (1999).
- Honda, A., Ueda, K., Nagata, K. & Ishihama, A. RNA polymerase of influenza virus: role of NP in RNA chain elongation. *J. Biochem.* **104**, 1021–1026 (1988).
- Shapiro, G. I. & Krug, R. M. Influenza virus RNA replication *in vitro*: synthesis of viral template RNAs and virion RNAs in the absence of an added primer. *J. Virol.* **62**, 2285–2290 (1988).
- Biswas, S. K., Boutz, P. L. & Nayak, D. P. Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *J. Virol.* **72**, 5493–5501 (1998).
- Gabriel, G. *et al.* The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl Acad. Sci. USA* **102**, 18590–18595 (2005).
- Otwiński, Z. & Minor, W. (eds) *Processing of X-ray Diffraction Data in Oscillation Mode* (Academic Press, New York, 1997).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
29. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
30. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Vikharia for providing the baculovirus for expressing nucleoprotein from influenza virus A/Hong Kong/1074/99 (H9N2); Z. Xie for electron microscopy analysis; P. Sliz for collecting the nucleoprotein native data; Y. Shamoo and G. Wu for providing the 24-nucleotide ssRNA; and

Y. Shamoo, D. Mata and S. Harrison for discussions and critical reading of the manuscript. Nucleoprotein diffraction data were collected at the Brookhaven National Synchrotron Light Source and the Cornell High Energy Synchrotron Source. This work is supported by grants from the Welch Foundation (to Y.J.T and R.M.K), National Institutes of Health (to Y.J.T), and the Nanoscale Science and Engineering Initiative of the National Science Foundation.

Author Contributions Q.Y. crystallized the nucleoprotein. Q.Y. and Y.J.T solved the nucleoprotein crystal structure. Y.J.T, Q.Y. and R.M.K. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The coordinates and the structure factor files have been deposited in the Protein Data Bank under accession code 2IQH. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.J.T (ytao@rice.edu).

Inhibition of Dll4 signalling inhibits tumour growth by deregulating angiogenesis

John Ridgway^{1*}, Gu Zhang^{1*}, Yan Wu^{2*}, Scott Stawicki², Wei-Ching Liang², Yvan Chanthery¹, Joe Kowalski¹, Ryan J. Watts¹, Christopher Callahan³, Ian Kasman³, Mallika Singh⁴, May Chien⁴, Christine Tan², Jo-Anne S. Hongo², Fred de Sauvage⁴, Greg Plowman¹ & Minhong Yan¹

Haploinsufficiency of Dll4, a vascular-specific Notch ligand, has shown that it is essential for embryonic vascular development and arteriogenesis^{1–3}. Mechanistically, it is unclear how the Dll4-mediated Notch pathway contributes to complex vascular processes that demand meticulous coordination of multiple signalling pathways. Here we show that Dll4-mediated Notch signalling has a unique role in regulating endothelial cell proliferation and differentiation. Neutralizing Dll4 with a Dll4-selective antibody rendered endothelial cells hyperproliferative, and caused defective cell fate specification or differentiation both *in vitro* and *in vivo*. In addition, blocking Dll4 inhibited tumour growth in several tumour models. Remarkably, antibodies against Dll4 and antibodies against vascular endothelial growth factor (VEGF) had paradoxically distinct effects on tumour vasculature. Our data also indicate that Dll4-mediated Notch signalling is crucial during active vascularization, but less important for normal vessel maintenance. Furthermore, unlike blocking Notch signalling globally, neutralizing Dll4 had no discernable impact on intestinal goblet cell differentiation^{4,5}, supporting the idea that Dll4-mediated Notch signalling is largely restricted to the vascular compartment. Therefore, targeting Dll4 might represent a broadly efficacious and well-tolerated approach for the treatment of solid tumours.

Recent genetic studies have provided convincing evidence that Dll4-mediated Notch signalling is essential for embryonic vascular development^{1–3}. However, it remains unclear how this signalling axis interacts with other pathways to ensure the formation of functional vessels. Preferential expression of Dll4 in tumour vasculature as opposed to normal tissue vessels indicates that it might have a role in tumour angiogenesis^{3,6}, but this is still speculative. We have generated neutralizing anti-Dll4 antibodies, to allow us to circumvent the embryonic lethality observed in Dll4 knockout studies^{1–3} and to explore the consequences of blocking Dll4 in postnatal life on normal or pathological vascular processes. One humanized phage antibody, YW152F, effectively blocked the interaction of Notch with Dll4, but not other Notch ligands^{7,8} (see Supplementary Fig. 1). Importantly, this antibody has similar dissociation constant (K_d) values against human and murine Dll4 (K_d of 0.15 and 0.09 nM, respectively), allowing us to evaluate it in mouse models.

Human umbilical vein endothelial cells (HUVECs) growing in fibrin gels in the presence of co-cultured human skin fibroblast cells generate sprouts with a distinct lumen-like structure⁹. Addition of YW152F markedly increased endothelial cell sprouting (Fig. 1a and Supplementary Fig. 2a, b). Proteolytic processing of Notch, catalysed by the γ -secretase activity of a protein complex, is an essential step in Notch activation¹⁰. Interestingly, the γ -secretase inhibitor dibenzazepine (DBZ)^{4,11} had the same effect on HUVEC sprouting. Given the

distinct mechanisms of these two treatments, the enhanced sprouting was clearly attributable to the attenuation of Notch signalling. Staining for the nuclear antigen Ki67, a well-established marker for proliferation, revealed that the enhanced endothelial cell sprouting was due to elevated cell proliferation (Fig. 1b and Supplementary Fig. 2c). In the original fibrin gel assay, HUVEC sprouting and subsequent lumen formation are supported by cocultured skin fibroblasts. When these cells were replaced with conditioned medium, both anti-Dll4 and DBZ could still enhance HUVEC sprouting (Fig. 1c), further supporting the idea that Dll4/Notch signalling has an autonomous role in endothelial cells. In the converse experiment, activation of Notch by immobilized Dll4 protein significantly inhibited growth (Fig. 1d). It has also been reported recently that Dll4 overexpression inhibits growth of endothelial cells¹². These findings indicate that the activation status of Dll4/Notch signalling is closely associated with endothelial cell proliferation.

Shortly after birth, the mouse retina develops a stereotypical vascular pattern in a well defined sequence of events^{13–15}. Prominent and dynamic expression of Dll4 in growing endothelial cells in the neonatal retinas indicates that Dll4 might help to regulate retinal vascular development¹⁶. Systemic delivery of YW152F to neonatal mice caused a profound alteration in the retinal vasculature. There was a massive accumulation of endothelial cells in the retina, generating a sheet-like structure with primitive vascular morphology (Fig. 1e and Supplementary Fig. 3a). A significant increase in Ki67 labeling in endothelial cells indicated that endothelial cell proliferation was elevated (Fig. 1f and Supplementary Fig. 3b). This hyperproliferative phenotype of retinal endothelial cells upon Dll4 blockade in neonatal mice corroborated our *in vitro* findings.

VEGF controls several fundamental aspects of endothelial cell function^{17,18}. It is unclear how VEGF signalling is integrated into complex vascular developmental processes, such as arteriovenous differentiation and hierarchical vascular organization, which evidently require additional coordinated signalling pathways. Genetic studies in zebra fish indicate that VEGF acts upstream of the Notch pathway during arterial endothelial differentiation¹⁹. We found that VEGF stimulation of HUVECs caused increased surface expression of Dll4 (data not shown), consistent with a recent report on the upregulation of Dll4 mRNA by VEGF stimulation⁶. Intriguingly, Dll4 itself is upregulated after Notch activation (see Supplementary Fig. 4), indicating that Dll4 uses a positive-feedback mechanism to relay VEGF signalling to the Notch pathway.

Notably, the hyperproliferation of endothelial cells that resulted from blocking Notch signalling still depended on VEGF. In the three-dimensional fibrin gel culture, anti-VEGF antibodies abolished most of the endothelial cell sprouting, with or without DBZ (Fig. 1g),

¹Tumor Biology & Angiogenesis, ²Antibody Engineering, ³Pathology and ⁴Molecular Biology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA.

*These authors contributed equally to this work.

raising the possibility that the hyperproliferation could be in part due to enhanced VEGF signalling. Indeed, blocking of Notch by YW152F or DBZ resulted in upregulation of VEGF receptor 2 (VEGFR2) (Fig. 1h). Conversely, activation of Notch by immobilized Dll4 suppressed the expression of VEGFR2 (Fig. 1h). Therefore, although VEGF can act upstream of the Dll4/Notch pathway, Dll4/Notch can fine-tune the response by negatively regulating VEGFR2 expression.

Besides the increase in endothelial cell proliferation, antagonizing Dll4/Notch caused a marked change in the morphology of endothelial cell sprouts in fibrin gel. The multicellular lumen-like structures were mostly absent (Fig. 2a), indicating defective differentiation of endothelial cells. Interestingly, Gale *et al.* also found a lack of lumens with nonorganized endothelial cell clusters in the aortas of Dll4^{+/-} embryos³. In YW152F-treated retinas, the characteristic pattern of radially alternating arteries and veins was severely disrupted. Staining with antibodies against α -smooth muscle actin (ASMA), which is associated with retinal arteries, was completely absent (Fig. 2b). This observation was remarkably similar to the defective arterial development in Dll4^{+/-} embryos¹⁻³. These findings highlight the essential role of Dll4/Notch in regulating the differentiation of endothelial cells.

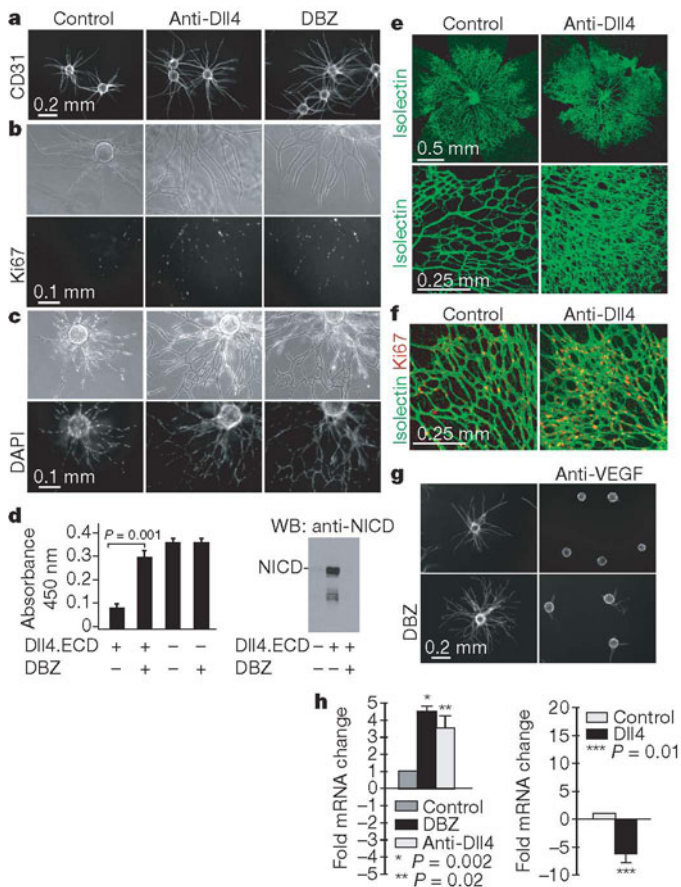


Figure 1 | Dll4-mediated Notch signalling regulates endothelial cell proliferation. Anti-Dll4 antibodies or DBZ caused enhanced sprouting (a) and proliferation (b) of HUVECs in co-culture with skin fibroblasts, or in the presence of skin fibroblast-conditioned medium (c). d, Immobilized Dll4 inhibited HUVEC proliferation (left) through Notch activation (right). WB, western blot. NICD, Notch intracellular domain. e, f, Systemic delivery of anti-Dll4 antibodies caused massive accumulation (e) and increased proliferation (f) of endothelial cells in neonatal retinas. g, Anti-VEGF antibodies inhibited HUVEC sprouting. h, Changes in VEGFR2 mRNA expression in response to Notch blockade in three-dimensional fibrin gel culture (left), or Notch activation by Dll4 in two-dimensional culture (right). Bars represent the mean $\pm$ s.e.m. ($n = 4$).

To address the possible role of Dll4/Notch signalling during tumour angiogenesis, we investigated the effect of blocking Dll4 on tumour growth in preclinical tumour models (Fig. 3a–f and Supplementary Fig. 6). In the HM7, Colo205 and Calu6 xenograft tumour models (Fig. 3a–c), we started treatment with YW152F after tumours had become established (≥ 250 mm³ tumour size). In all three models, the control and treatment groups showed differences in tumour growth rates three days after dosing. The tumour volume of the treatment group remained static over two weeks of treatment. In addition to subcutaneous tumours, anti-Dll4 antibodies also inhibited tumours growing in mouse mammary fat pads. In the MDA-MB-435 tumour model, treatment was started 14 days after injection of tumour cells. There was a difference in tumour growth curves between the control and treatment groups within six days of dosing, which became increasingly significant as treatment continued (Fig. 3d).

In light of the inhibition of tumour growth, we used the EL4 mouse lymphoma tumour model for vascular histology studies. We found that treatment with anti-Dll4 antibodies resulted in a marked increase in endothelial cell density (Fig. 3g). Treatment had similar effects in HM7, Calu6 and MDA-MB-435 tumours (see Supplementary Fig. 5). By contrast, anti-VEGF antibodies had the opposite effect (Fig. 3g), although both treatments were similarly effective in the EL4 model (see Supplementary Fig. 6a).

As *in vitro* blocking of the Dll4/Notch pathway impaired the formation of lumen-like structures by endothelial cells (Fig. 2a), anti-Dll4 antibodies could cause a similar defect in the tumour vasculature and affect efficient blood flow. Systemic perfusion with fluorescein isothiocyanate (FITC)-conjugated lectin revealed that anti-Dll4 treatment resulted in a marked reduction in lectin labelling of tumour vessels (Fig. 3h). Given the essential role of Dll4/Notch signalling in arteriovenous differentiation, both in embryos and in early postnatal retinas, anti-Dll4 antibodies could effect the cell fate specification of tumour endothelial cells and cause defective directional blood flow. Indeed, in anti-Dll4-treated Colo205 tumours, there were regions where high endothelial cell density was associated with low viable tumour cell content, consistent with poor vascular function (Supplementary Fig. 7). Further studies using vascular imaging techniques are needed to gain insights into the precise vascular defects in these treated tumours.

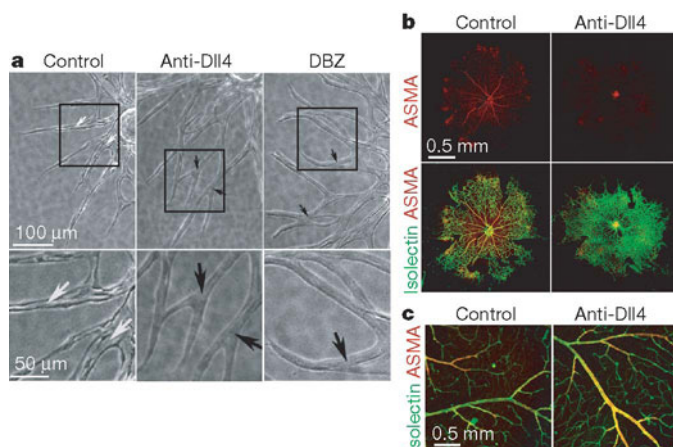


Figure 2 | Dll4-mediated Notch signalling regulates endothelial cell differentiation. a, The lumen-like structures (white arrows) formed by HUVECs growing in fibrin gels were absent in the presence of anti-Dll4 antibodies or DBZ. Instead, the sprouts were highly packed with cells (black arrows). The boxed area is shown enlarged in the bottom micrograph. b, Anti-Dll4 antibodies block arterial development. Confocal images of neonatal mouse retinas stained with α -smooth muscle actin (ASMA; red) and islectin (green). c, Confocal images of adult mouse retinas stained with ASMA and islectin. Eight-week-old mice were treated with PBS or anti-Dll4 antibodies (10 mg kg⁻¹, twice weekly) for two weeks.

In tumour models, anti-Dll4 and anti-VEGF antibodies had opposite effects on tumour vasculature (Fig. 3g), indicating that they have non-overlapping mechanisms of action. Although approved therapies targeting VEGF signalling are showing expanding clinical benefits for the treatment of cancer^{20,21}, tremendous challenges are anticipated when resistance to or escape from such therapies evolve. In the Calu6 xenograft model, anti-VEGF antibodies potently inhibited tumour growth (see Supplementary Fig. 6b). By comparison, the response of MV-522 xenograft tumours to anti-VEGF antibodies was much reduced. In this model, anti-Dll4 antibodies were also only modestly effective at reducing tumour growth. Importantly, combination therapy with both antibodies significantly enhanced the inhibition of tumour growth (Fig. 3e). Interestingly, WEHI3 mouse tumours were completely unresponsive to the same anti-VEGF treatment as was used in the Calu6 and MV-522 tumour studies (Fig. 3f).

However, treatment with anti-Dll4 antibodies resulted in significant inhibition of tumour growth (Fig. 3f). Therefore, Dll4 inhibition might represent a promising strategy for treating tumours in which anti-VEGF treatment is less effective.

An important concern about global inhibition of Notch is that it might be associated with potential side effects, given the pleiotropic roles of Notch signalling in regulating the homeostasis of postnatal self-renewal systems. For instance, Notch signalling is required to maintain undifferentiated crypt progenitor cells in the intestines^{4,5}. Inhibitors of γ -secretase, which indiscriminately block all Notch activities, cause unwanted side effects in rodents owing to a massive increase in goblet cells in the crypt compartment^{11,22}. We used immunohistochemistry to examine the small intestines of mice treated with anti-Dll4 antibodies. Unlike mice treated with DBZ, these mice showed no differences in epithelial crypt cell differentiation or

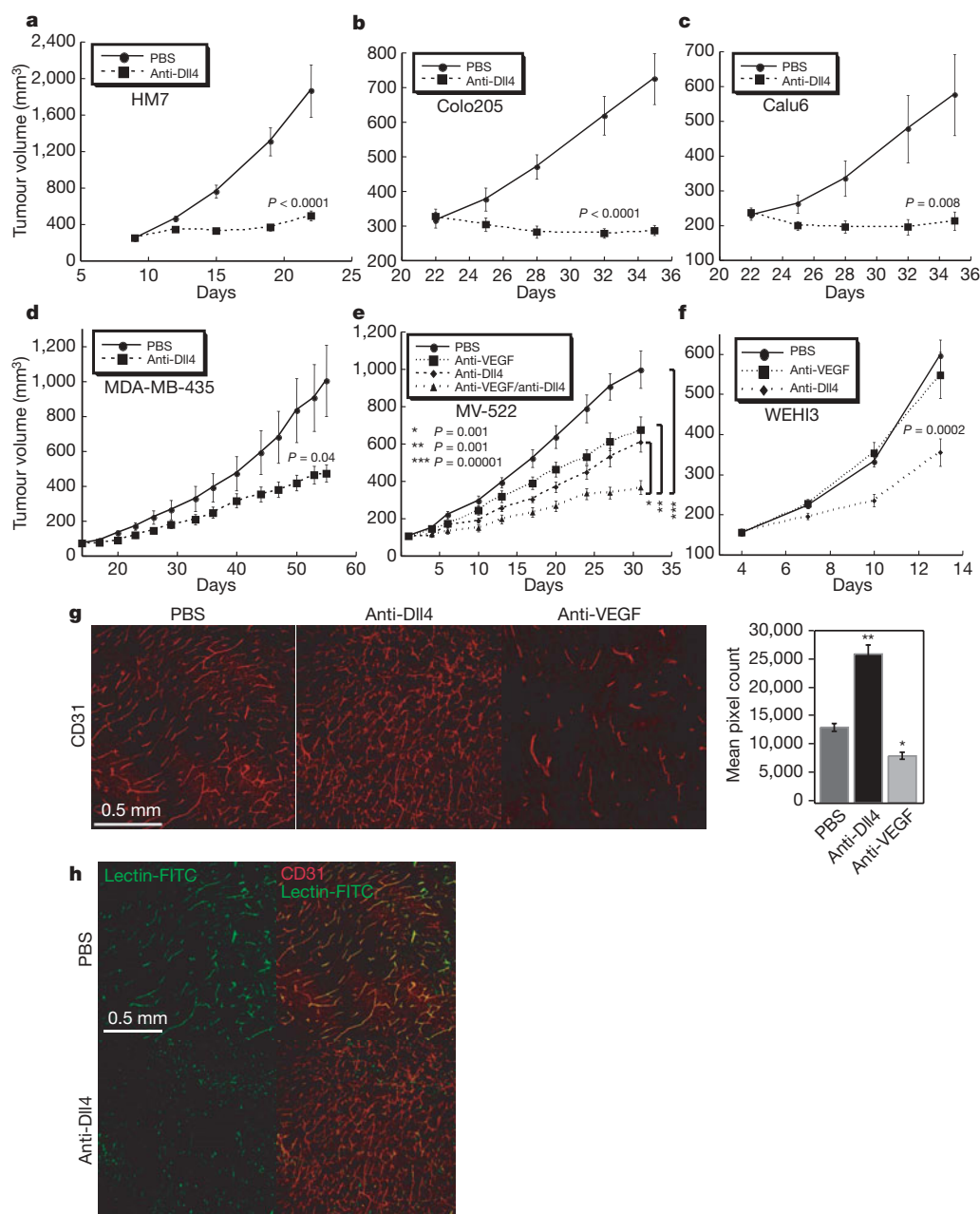


Figure 3 | Selective blocking of Dll4 disrupts tumour angiogenesis and inhibits tumour growth. **a–f**, Results of six tumour models: HM7 (**a**), Colo205 (**b**), Calu6 (**c**), MDA-MB-435 (**d**), MV-522 (**e**) and WEHI3 (**f**). Mean tumour volumes with standard errors are presented ($n = 10$). **g, h**, Tumour vascular histology studies. **g**, Immunohistochemistry of

anti-CD31 in EL4 tumour sections from control, anti-Dll4 and anti-VEGF treated mice. Quantification of CD31 staining was performed using Image J. Bars represent the mean $\pm$ s.e.m. ($n = 6–10$, * and ** denote P values of 0.001 and 0.0001, respectively). **h**, Lectin perfusion and anti-CD31 staining of EL4 tumour sections.

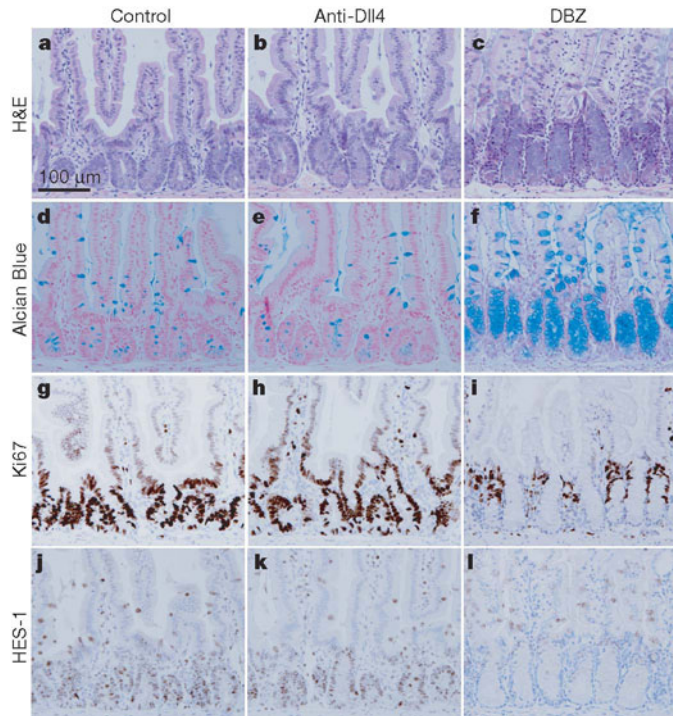


Figure 4 | Dll4/Notch in the homeostasis of mouse small intestine.

Immunohistochemical studies of small intestines from control (a, d, g, j), anti-Dll4 (b, e, h, k), and DBZ-treated (c, f, i, l) mice. As shown by haematoxylin and eosin (H&E; a–c) and Alcian Blue staining (d–f), DBZ treatment, but not anti-Dll4 antibodies, resulted in conversion of the transit-amplifying crypt cell population to goblet cells. Ki67 staining (g–i) showed similar proliferation rates for control and anti-Dll4-treated animals (~75%), whereas DBZ-treated groups showed <10% proliferation. Staining with antibodies against HES-1 produced 70% positivity for control and anti-Dll4-treated animals, whereas DBZ-treated animals showed <2% positive (j–l).

proliferation when compared with control groups after six weeks of treatment (Figs. 4a–i). Furthermore, anti-Dll4 antibodies did not alter the expression of the Notch target gene HES-1 in the transit-amplifying compartments (Figs. 4j–l). These results support the idea that Dll4/Notch signalling is largely restricted to the vascular system.

Although blocking Dll4 had a profound influence on retinal vascular development in neonatal mice, administration of anti-Dll4 antibodies has no visible effect on adult retinal vasculature (Fig. 2c). Thus, Dll4/Notch signalling is crucial during active angiogenesis, but less important for normal vessel maintenance. Consistent with this, during the course of anti-Dll4 treatment, we saw no apparent weight loss or animal death in tumour-bearing mice when dosed at 10 mg kg⁻¹ twice per week for up to 8 weeks (data not shown). Therefore, targeting Dll4 might be a broadly efficacious and well tolerated approach for the treatment of solid tumours.

METHODS

HUVEC fibrin gel bead assay. Details of the HUVEC fibrin gel bead assay have been described⁹. Briefly, Cytodex 3 beads (Amersham Pharmacia Biotech) were coated with 350–400 HUVECs per bead. About 200 HUVEC-coated beads were imbedded in a fibrin clot in one well of a 12-well tissue culture plate. 8 × 10⁴ skin fibroblast cells were plated on top of the clot. Assays were terminated between day 7 and day 9 for immunostaining and imaging. In some experiments, HUVEC sprouts were visualized by staining with Biotin-anti-CD31 (clone WM59, eBioscience) and streptavidin-Cy3. For HUVEC nucleus staining, fibrin gels were fixed overnight in 2% paraformaldehyde (PFA), and stained with 4', 6-diamidino-2-phenylindole (DAPI, Sigma). For Ki67 staining, fibrin gels were treated with 10 × trypsin–EDTA for 5 min to remove the top layer of skin fibroblasts, neutralized with 10% FBS in PBS, and fixed overnight in 4% PFA. Fibrin gels were then blocked with 10% goat serum in PBST (PBS, 1% BSA, 0.5% Triton X-100) for 4 h, incubated overnight with rabbit anti-mouse Ki67 (Ready-To-Use, clone Sp6, LabVision), followed by secondary detection with anti-rabbit

IgG-Cy3 (Jackson ImmunoResearch). All overnight incubations were done at 4 °C. In all experiments, anti-Dll4 antibodies and DBZ were used at 5 µg ml⁻¹ and 0.08 µM, respectively.

Tumour vascular labelling and immunohistochemistry. Mice were anaesthetized with Isoflurane. FITC-labelled lycopersicon esculentum lectin (150 µg in 150 µl of 0.9% NaCl; Vector Laboratories) was injected intravenously and allowed to circulate for 5 min before systemic perfusion. The vasculature was perfused transcardially with 1% PFA in PBS for 3 min. Tumours were removed and post fixed by immersion in the same fixative for 2 h, followed by an incubation in 30% sucrose overnight for cryoprotection, then embedded in OCT (Optimal Cutting Temperature) compound (Sakura Finetek USA). Sections (40 µm thickness) were stained with anti-mouse CD31 (1:50, BD Pharmingen), followed by Alexa 594 goat anti-rat IgG (1:800, Molecular Probes).

Histology and immunohistochemistry of mouse intestines. Small intestines were dissected from control mice, mice treated with anti-Dll4 antibodies (10 mg kg⁻¹, twice weekly for 6 weeks), or mice treated with DBZ (30 µmol kg⁻¹ daily for 5 days). Formalin-fixed and paraffin-embedded mouse small intestine tissues were sectioned at 3 µm thickness. Histochemical identification of intestinal cell types was performed with Alcian blue as recommended by the manufacturer (PolyScientific). For anti-Ki67 staining, sections were pretreated with Target Retrieval Solution (S1700, DAKO), incubated with rabbit anti-Ki67 (1:200, clone SP6, Neomarkers), followed by a goat anti-rabbit (7.5 µg/ml, Vector Laboratories) and stained with the Vectastain ABC Elite Kit (Vector Laboratories). For HES-1 staining, sections were pretreated with Target Retrieval Solution (S1700, DAKO), incubated with an anti-rat HES-1 antibody (1 µg ml⁻¹, clone NM1, MBL International), followed by a biotinylated rabbit anti-rat secondary (2.5 µg ml⁻¹) and TSA–HRP Kit (T20943, Invitrogen). All immunoperoxidase-stained sections were counterstained with Mayer's haematoxylin. Quantification of nuclear Ki67 and HES-1 staining was based on the assessment of 300 jejunum crypt cells per animal.

Received 22 May; accepted 2 October 2006.

- Duarte, A. *et al.* Dosage-sensitive requirement for mouse Dll4 in artery development. *Genes Dev.* **18**, 2474–2478 (2004).
- Krebs, L. T. *et al.* Haploinsufficient lethality and formation of arteriovenous malformations in Notch pathway mutants. *Genes Dev.* **18**, 2469–2473 (2004).
- Gale, N. W. *et al.* Haploinsufficiency of delta-like 4 ligand results in embryonic lethality due to major defects in arterial and vascular development. *Proc. Natl Acad. Sci. USA* **101**, 15949–15954 (2004).
- van Es, J. H. *et al.* Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**, 959–963 (2005).
- Fre, S. *et al.* Notch signals control the fate of immature progenitor cells in the intestine. *Nature* **435**, 964–968 (2005).
- Patel, N. S. *et al.* Up-regulation of delta-like 4 ligand in human tumor vasculature and the role of basal expression in endothelial cell function. *Cancer Res.* **65**, 8690–8697 (2005).
- Lee, C. V. *et al.* High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. *J. Mol. Biol.* **340**, 1073–1093 (2004).
- Liang, W. C. *et al.* Cross-species vascular endothelial growth factor (VEGF)-blocking antibodies completely inhibit the growth of human tumor xenografts and measure the contribution of stromal VEGF. *J. Biol. Chem.* **281**, 951–961 (2006).
- Nakatsu, M. N. *et al.* Angiogenic sprouting and capillary lumen formation modeled by human umbilical vein endothelial cells (HUVEC) in fibrin gels: the role of fibroblasts and Angiopoietin-1. *Microvasc. Res.* **66**, 102–112 (2003).
- Baron, M. An overview of the Notch signalling pathway. *Semin. Cell Dev. Biol.* **14**, 113–119 (2003).
- Milano, J. *et al.* Modulation of notch processing by gamma-secretase inhibitors causes intestinal goblet cell metaplasia and induction of genes known to specify gut secretory lineage differentiation. *Toxicol. Sci.* **82**, 341–358 (2004).
- Williams, C. K., Li, J. L., Murga, M., Harris, A. L. & Tosato, G. Up-regulation of the Notch ligand Delta-like 4 inhibits VEGF-induced endothelial cell function. *Blood* **107**, 931–939 (2006).
- Stone, J. & Dreher, Z. Relationship between astrocytes, ganglion cells and vasculature of the retina. *J. Comp. Neurol.* **255**, 35–49 (1987).
- Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
- Fruttiger, M. Development of the mouse retinal vasculature: angiogenesis versus vasculogenesis. *Invest. Ophthalmol. Vis. Sci.* **43**, 522–527 (2002).
- Claxton, S. & Fruttiger, M. Periodic Delta-like 4 expression in developing retinal arteries. *Gene Expr. Patterns* **5**, 123–127 (2004).
- Ferrara, N. The role of VEGF in the regulation of physiological and pathological angiogenesis. *EXS* **94**, 209–231 (2005).
- Coults, L., Chawengsaksohak, K. & Rossant, J. Endothelial cells and VEGF in vascular development. *Nature* **438**, 937–945 (2005).
- Lawson, N. D. *et al.* Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* **128**, 3675–3683 (2001).
- Ferrara, N. & Kerbel, R. S. Angiogenesis as a therapeutic target. *Nature* **438**, 967–974 (2005).

21. Jain, R. K., Duda, D. G., Clark, J. W. & Loeffler, J. S. Lessons from phase III clinical trials on anti-VEGF therapy for cancer. *Nature Clin. Pract. Oncol.* **3**, 24–40 (2006).
22. Wong, G. T. *et al.* Chronic treatment with the gamma-secretase inhibitor LY-411,575 inhibits beta-amyloid peptide production and alters lymphopoiesis and intestinal cell differentiation. *J. Biol. Chem.* **279**, 12876–12882 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Fuh and V. Lee for the antibody phage libraries, L. DeGuzman and J. Zavala-Solorio for technical assistance, M. Vasser and P. Ng for oligonucleotide synthesis and purification, Genentech Protein Chemistry, Pathology Lab and core DNA sequencing facility for support services, W. Ye, C. Siebel, N. Ferrara, J. Zha and K. O'Rourke for discussion, and V. Dixit for encouragement.

Author Contributions J. R. and G. Z. designed and performed experiments. Y. W. supervised antibody generation. S. S. and W. L. generated antibodies. Y. C., J. K. and R. J. W. performed neonatal mouse retina studies and data analysis. C. C. performed pathological analysis. I. K. provided expertise and assistance on confocal imaging. M. S. and M. C. performed mouse intestine studies. C. T. and J. H. generated analytical reagents. F. de S. supervised mouse intestine studies. G. P. supervised tumour studies of combination therapy. M. Y. designed and conceptualized the study, and wrote the manuscript. All authors made comments on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.Y. (minhong@gene.com).

LETTERS

Essential role for collectrin in renal amino acid transport

Ursula Danilczyk¹, Renu Sarao¹, Christine Remy², Chahira Benabbas², Gerti Stange², Andreas Richter³, Sudha Arya⁴, J. Andrew Pospisilik¹, Dustin Singer², Simone M. R. Camargo², Victoria Makrides², Tamara Ramadan², Francois Verrey², Carsten A. Wagner² & Josef M. Penninger¹

Angiotensin-converting enzyme 2 (ACE2) is a regulator of the renin-angiotensin system involved in acute lung failure, cardiovascular functions and severe acute respiratory syndrome (SARS) infections in mammals^{1–3}. A gene encoding a homologue to ACE2, termed collectrin (Tmem27), has been identified in immediate proximity to the *ace2* locus⁴. The *in vivo* function of collectrin was unclear. Here we report that targeted disruption of *collectrin* in mice results in a severe defect in renal amino acid uptake owing to downregulation of apical amino acid transporters in the kidney. Collectrin associates with multiple apical transporters and defines a novel group of renal amino acid transporters. Expression of collectrin in *Xenopus* oocytes and Madin–Darby canine kidney (MDCK) cells enhances amino acid transport by the transporter B⁰AT1. These data identify collectrin as a key regulator of renal amino acid uptake.

ACE2 is a chimaeric protein that has emerged from the duplication of two genes: homology with ACE at the catalytic domain and homology with collectrin in the membrane proximal domain (Fig. 1a). In the mouse, *collectrin* is expressed in the kidneys⁴ and pancreas^{5,6} and to a lesser extent in the intestine (duodenum, jejunum and ileum), liver, heart and stomach (Supplementary Fig. 1). To assess the *in vivo* function of collectrin, we generated *collectrin* knockout mice (Fig. 1b). Mice were crossed with the transgenic Cre *deleter* strain to ubiquitously delete *collectrin*⁷ (Supplementary Fig. 2a). The null mutation of *collectrin* was verified by the absence of *collectrin* messenger RNA transcripts and protein (Supplementary Fig. 2b, c). ACE2 expression in the kidneys was not altered in *collectrin* mutant mice (Supplementary Fig. 2c). *Collectrin* null mice were born at the expected mendelian frequency, indistinguishable from their heterozygous and wild-type littermates. *Collectrin* mutant males (*collectrin*^{−/y}) and females (*collectrin*^{−/−}) were fertile and we failed to observe any overt morphological alterations in all organs analysed, including kidney, for up to six months of age (Fig. 1c).

Collectrin was originally reported to be expressed in collecting ducts⁴ and β -islet cells of the pancreas^{5,6}. We therefore assessed whether collectrin is involved in renal salt and/or glucose balance. Na⁺ excretion in urine was comparable in *collectrin*^{+/y} and *collectrin*^{−/y} mice. Loss of collectrin also did not affect the levels of calcium, potassium, phosphate, chloride, urea, uric acid or creatinine in urine and blood (Supplementary Table 1). *collectrin* mutant mice had the same glucose concentration as their wild-type littermates in serum and urine. Surprisingly, storage of the urine samples collected from mutant mice at 4 °C resulted in the formation of large white precipitates (Fig. 1d). Morphological analysis revealed large needle-like crystals in the urine of the *collectrin* null mice whereas such crystals were never observed in wild-type mice (Fig. 1e). Biochemically, these

crystals were not composed of common urinary crystal components such as ammonium urate, sodium urate, calcium oxalate dihydrate or calcium phosphate. High performance liquid chromatography (HPLC) analyses showed that these crystals contained ~10% phenylalanine (Phe) and 90% tyrosine (Tyr) (see Fig. 1f). When animals were fed a tyrosine-free and phenylalanine-free diet or when the animals were fasting, no crystals were formed in the urine of *collectrin* null mice (Fig. 1d), indicating that the observed changes are not caused by tyrosinaemia^{8,9}. Thus, loss of collectrin expression results in the formation of tyrosine/phenylalanine crystals in the urine.

To determine whether collectrin deficiency caused a more general defect in amino acid handling in the kidneys, we assessed urinary amino acid contents. Whereas in wild-type mice almost all amino acids are reabsorbed from the urine, deletion of *collectrin* resulted in severe leakage of all amino acids (Fig. 1g). Concentrations of the amino acids were also lower in the serum of mutant mice, although amino acids were still present (Supplementary Fig. 3). Urinary output was higher in *collectrin*^{−/y} mice (3.31 ± 0.25 ml urine per 24 h) as compared to *collectrin*^{+/y} littermates (1.65 ± 0.28 ml urine per 24 h) which may be owing to the higher concentration of osmotically active solutes in urine from *collectrin*-deficient mice. Water restriction for 24 h demonstrated that *collectrin*-deficient mice were not able to concentrate urine to the same degree, confirming that the higher urinary output is due to a primarily renal defect (Supplementary Fig. 4). Thus, collectrin is critical for normal amino acid reabsorption in the kidney.

Amino acid reabsorption occurs in the proximal tubules of the kidney¹⁰. Immunohistochemical staining showed that collectrin is indeed expressed at the luminal side of brush border membranes in proximal tubules (Fig. 2a), which is a specialized microenvironment involved in amino acid reabsorption¹¹. Biochemically, we confirmed specific expression of collectrin in brush border membrane vesicles (BBMV) isolated from the kidney of *collectrin*^{+/y}, but not *collectrin*^{−/y} mice (Fig. 2b). BBMV isolates were free from contamination (assessed by the absence of the collecting duct marker aquaporin 2; Fig. 2b). In addition, collectrin mRNA was found to be highly expressed in proximal tubules (Fig. 2c). Thus, collectrin is localized in brush border membranes of proximal tubules.

Because collectrin is predicted to be a type I transmembrane protein having only one transmembrane helix⁴, it was unlikely that collectrin itself forms an amino acid transport pore. Because collectrin expression parallels that of the B⁰ cluster of the solute carrier SLC6 proteins B⁰AT1, XT3s1/SIT1, XT2 and XT3, a recently identified subfamily of Na⁺-dependent transporters of neutral amino acids^{12–17}, we assessed whether loss of collectrin may affect these

¹IMBA, Institute for Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr-Gasse 3, A-1030, Vienna, Austria. ²Institute of Physiology and Center for Integrative Human Physiology, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. ³Department of Chemical Ecology and Ecosystem Research, Center of Ecology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria. ⁴The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada.

amino acid transporters. Immunoblots of BBMVs revealed that the neutral amino acid transporter B⁰AT1 was markedly downregulated in *collectrin*^{-/-} mice (Fig. 2d and Supplementary Fig. 5a, b). Protein expression of the three other members of the SLC6 family XT3s1/SIT1, XT2 and XT3, and expression of EAAC1, the apical transporter responsible for glutamate/aspartate reabsorption¹⁸, was also markedly downregulated (Fig. 2e). By contrast, expression of b^{0,+}AT (also known as slc7a9) (Fig. 2e), the exchanger mediating apical uptake of cystine, arginine, lysine and ornithine¹⁹, and expression of the basolateral amino acid transporter subunits 4F2hc and LAT2 (ref. 10) remained unchanged (Supplementary Fig. 5b). Expression of the sodium/phosphate co-transporter Na⁺/Pi IIa was also not affected in *collectrin* null mice (Fig. 2d). mRNA expression of all these transporters was comparable in *collectrin*^{+/y} and *collectrin*^{-/-} mice (Supplementary Fig. 6). Immunoblotting of total kidney membranes demonstrated that the abundance of these amino acid transporters was also decreased to a similar extent as observed in BBMVs (Supplementary Fig. 5b).

To test whether reduced expression resulted in decreased amino acid transport activity, we isolated BBMVs and measured Na⁺-dependent amino acid uptake rates. As expected from the urinary

loss of B⁰AT1 substrates, the uptake rate of glutamine and phenylalanine in BBMVs from kidneys of *collectrin* null mice was decreased compared to control littermates, whereas uptake of phosphate was not affected (Fig. 2f). These data show that loss of *collectrin* is associated with reduced expression and impaired activity of apical, Na⁺-dependent amino acid transporters in the kidney.

To identify the molecular pathway by which *collectrin* regulates amino acid transporters, we first examined whether *collectrin* can associate with B⁰AT1 and/or other transporters. Immunoprecipitations from BBMVs showed direct binding between *collectrin* and B⁰AT1, XT2 and XT3 in *collectrin*^{+/y}, but not *collectrin*^{-/-} mice (Fig. 3a). Moreover, *collectrin* co-localized with B⁰AT1 in the early proximal S1 tubule (Fig. 3b). As a control, *collectrin* did not associate with b^{0,+}AT (Fig. 3a). The specificity of the *collectrin* band was confirmed by pre-incubation with the antigenic peptide (Supplementary Fig. 7a, b). Renal amino acid transporters such as LAT2 or γ⁺LAT1 form functional heterodimers with a second subunit, 4F2hc (ref. 10). In all known instances, this heterodimerization is stabilized by disulphide bonds²⁰. Under reducing conditions (+DTT), *collectrin* appears as a single band with gel mobility of ~44 kDa, whereas B⁰AT1 can be detected as a single band that corresponds to a monomeric form of

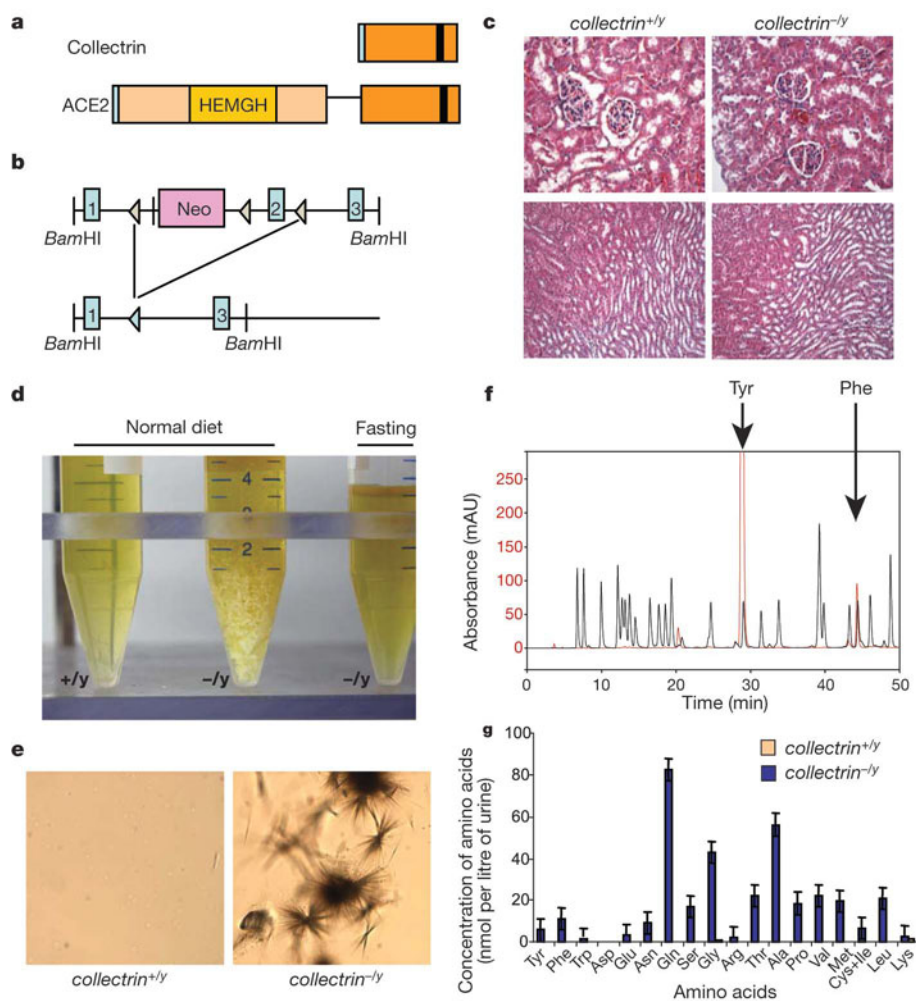


Figure 1 | Tyrosine crystals and renal amino acid loss in *collectrin*^{-/-} mice.

a, Schematic outline of homologous regions between collectrin and ACE2. Collectrin lacks the HEMGH carboxypeptidase domain. The black boxes indicate the transmembrane domains. Leader peptides are shown in light blue. **b**, Gene targeting (floxed-and-delete) strategy for the generation of *collectrin* mutant alleles. Upon breeding with Cre deleter mice, animals with excision of exon 2 were selected. **c**, H&E-stained sections of kidney isolated from six-month-old *collectrin*^{+/y} and *collectrin*^{-/-} mice. Top panels, overview over cortex; bottom panels, normal medullary structures. **d**, At

4 °C, large white precipitates form in the urine of *collectrin* null (-/-) but not control wild-type (+/y) mice. These precipitates disappear upon fasting for 24 h. **e**, Formation of needle-like crystals in the urine of *collectrin* null mice. **f**, HPLC analysis of crystals isolated from the urine of *collectrin*^{-/-} mice. The crystals co-elute with tyrosine and phenylalanine. mAU, milli-absorbance units. **g**, Excretion of all amino acids tested in *collectrin*^{-/-} mice. Urine was collected from four-month-old *collectrin*^{+/y} and *collectrin*^{-/-} littermates and analysed by HPLC. Values are mean ± s.e.m. of ten mice per group. *P* < 0.05 (Student's *t*-test and analysis of variance, ANOVA).

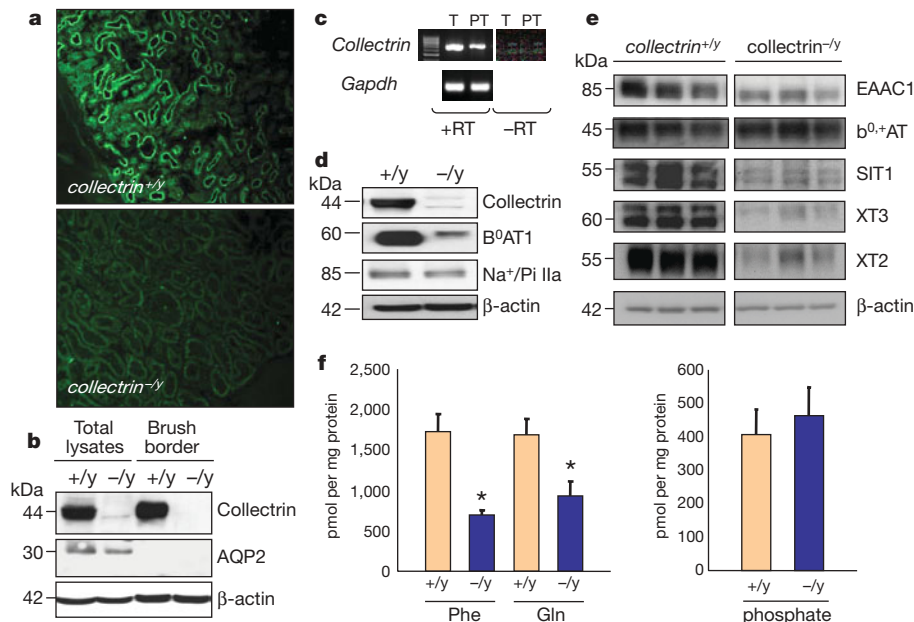


Figure 2 | Defective expression of amino acid transporters in *collectrin* mutant mice. **a**, Immunodetection of collectrin (green) in kidneys of *collectrin*^{+y/y} and *collectrin*^{-y/y} mice. Note that collectrin localizes to proximal tubules. **b**, Western blot analysis of brush border membranes from *collectrin*^{+y/y} and *collectrin*^{-y/y} mice. Protein extracts from total lysates (50 µg per lane) and renal brush border membranes (10 µg per lane) were analysed using antibodies to collectrin, aquaporin (AQP2), and β-actin. **c**, Collectrin mRNA expression in proximal tubules (PT) and total kidneys (T). RNA was isolated from wild-type mice and reverse transcribed (+RT). Equal loading

was confirmed by Gapdh. **d**, **e**, Western blot analysis of proteins (50 µg) from brush border membranes isolated from *collectrin*^{+y/y} and *collectrin*^{-y/y} mice. In **d** and **e**, data from three different mice are shown for each genotype. Molecular sizes are indicated in kDa. **f**, Na⁺-dependent uptake of the amino acids phenylalanine (Phe) and glutamine (Gln) and uptake of phosphate into BBMVs from *collectrin*^{+y/y} and *collectrin*^{-y/y} mice. All experiments were done in triplicates from each kidney. Values are mean ± s.e.m. of four mice per group. **P* < 0.001 (Student's *t*-test).

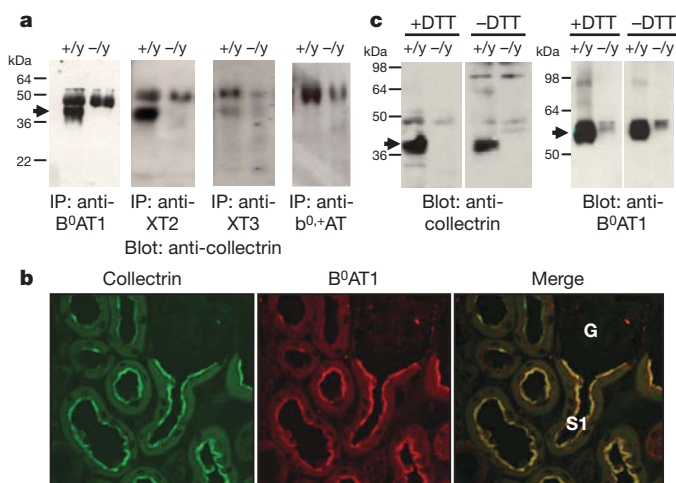


Figure 3 | Collectrin associates with apical amino acid transporters through non-covalent interactions. **a**, Renal brush border membrane proteins (100 µg) from *collectrin*^{+y/y} and *collectrin*^{-y/y} mice were incubated overnight with anti-B⁰AT1, anti-XT2, and anti-XT3 and anti-b⁰,+AT antibodies. Immunoprecipitated (IP) complexes were analysed by western blot using an anti-collectrin antibody (arrow). **b**, Collectrin co-localizes with B⁰AT1 in kidney proximal tubules. Kidney sections from wild-type mice were stained with antibodies to collectrin (green) and B⁰AT1 (red). Yellow indicates co-localization. S1, initial part of the proximal tubule. G, glomerulus. Magnifications, ×400. **c**, Western blot analysis of isolated renal brush border membranes under reducing (+DTT; 100 mM) and non-reducing (-DTT) conditions from *collectrin*^{+y/y} and *collectrin*^{-y/y} mice using antibodies directed against collectrin and B⁰AT1. Under reducing as well as under non-reducing conditions collectrin and B⁰AT1 appear as bands of ~44 and ~60 kDa, respectively (arrows). Molecular sizes are indicated in kDa.

B⁰AT1 (Fig. 3c). Unlike 4F2hc and its catalytic subunits, B⁰AT1 and collectrin migrate under non-reducing (no dithiothreitol, -DTT) conditions with the same mobility as under reducing conditions (Fig. 3c), indicating that the association between collectrin and B⁰AT1 as well as other transporters occurs through non-covalent interactions. Thus, collectrin co-localizes and specifically associates with a subgroup of renal amino acid transporters.

To define the molecular mechanism of collectrin-regulated amino acid transport, we examined the effect of collectrin on B⁰AT1 function using *Xenopus laevis* oocytes and polarized MDCK cells. Two days of expression of B⁰AT1 alone in *Xenopus* oocytes conferred only a low transport activity for the B⁰AT1 substrate L-isoleucine¹⁶ (Fig. 4a). Collectrin alone did not exhibit any transport activity. However, co-expression of B⁰AT1 with collectrin resulted in a striking increase in L-isoleucine uptake (Fig. 4a). As a control, co-expression of collectrin with the Na⁺/phosphate IIa co-transporter in oocytes did not affect the rates and kinetics of phosphate uptake (Fig. 4b). To assess whether collectrin changes binding of L-isoleucine to its transporter B⁰AT1, we measured the apparent affinity (*K_m*) for L-isoleucine uptake (Fig. 4c). Co-transport of L-isoleucine with Na⁺ generated a saturable, reversible inward current with an apparent *K_m* for L-isoleucine of 0.99 ± 0.17 mM when B⁰AT1 was expressed alone, and of 0.78 ± 0.10 mM when B⁰AT1 was co-expressed with collectrin. Thus, the collectrin-induced increase in L-isoleucine transport is not due to a change in apparent affinity, but due to an increase in the maximal transport rate of B⁰AT1, probably as a result of increased surface expression. Indeed, expression of collectrin allowed functional surface expression of B⁰AT1 and apical uptake of L-isoleucine in polarized kidney MDCK cells (Fig. 4d). Thus, collectrin enhances functional surface expression of the SLC6 family transporter B⁰AT1.

We conclude that genetic inactivation of *collectrin* in mice results in a major defect in renal amino acid reabsorption. Collectrin controls protein expression and function of apical amino acid transporters in the brush border membranes of proximal tubules.

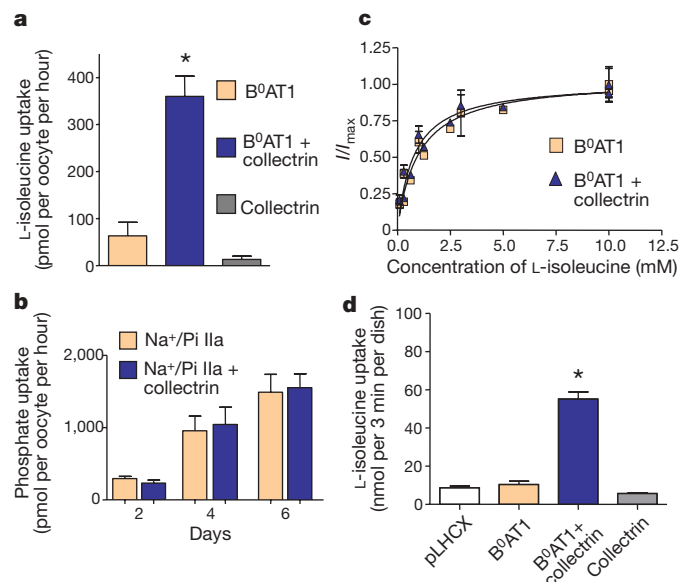


Figure 4 | Collectrin enhances B⁰AT1 transport activity. **a**, *Xenopus laevis* oocytes were injected with murine B⁰AT1, collectrin, or B⁰AT1 plus collectrin RNA. L-isoleucine transport was determined two days after injection. Each bar represents mean transport activity $\pm$ s.d. ($n = 15$ oocytes analysed). **b**, *Xenopus laevis* oocytes were injected with murine Na⁺/Pi IIa alone or Na⁺/Pi IIa plus collectrin; phosphate transport (mean $\pm$ s.d.) was then determined at the indicated time points. $n = 7$ oocytes per sample. **c**, The half-maximal uptake rate K_m of L-isoleucine by B⁰AT1 at six days of expression in the absence or presence of collectrin. K_m was determined using two electrode voltage clamps at -50 mV using 0–10 mM of L-isoleucine in the presence of 100 mM Na⁺. Data was normalized to 10 mM L-isoleucine for $n = 11$ oocytes per group. **d**, MDCK cells were transduced with murine B⁰AT1, collectrin, B⁰AT1 plus collectrin, or empty vector (pLHCX). Bars represent the mean uptake of 300 μ M L-isoleucine in 3 min $\pm$ s.d. * $P < 0.001$ (Student's t -test).

Mechanistically, collectrin associates with several amino acid transporters, enhances their surface expression, and thereby increases amino acid transport function. These data identify the ACE2 homologue collectrin as a regulator of renal amino acid uptake and may provide a molecular explanation for aminoaciduria in Hartnup disease, Fanconi syndrome, or diabetes.

METHODS

collectrin knockout mice. A gene-targeting vector containing three *loxP* sites flanking exon 2 and a Neo selection cassette was constructed. Chimaeric progeny were crossed to the C57BL/6J Cre *deleter* line⁷ to generate complete null animals. **mRNA and protein analyses.** Real-time polymerase chain reaction (PCR) was performed as described¹². For protein analyses, total cell lysates and BBMVs of adult mouse kidney were prepared²¹ and probed with antibodies to mouse b⁰+AT, NaPi IIa, aquaporin 2, collectrin, XT2, XT3, SIT1, B⁰AT1, ACE2, LAT2 and 4F2. For immunoprecipitations, brush border membrane proteins were probed with antibodies to B⁰AT1, XT2, XT3 and b⁰+AT.

Urine and plasma analyses. Urine was collected over 24 h in a metabolic cage. Urinary crystals were observed under a magnification of $\times 1,000$. For HPLC analysis¹⁹, crystals were dissolved in 0.1 M HCl. Na⁺, K⁺ and Cl⁻ were assayed by indirect potentiometry. Phosphate was determined using the phosphomolybdate method. Calcium levels were obtained by cresolphthalein. Glucose, urea and uric acid were measured using Reflovet (Roche).

Histology and immunohistochemistry. Paraffin sections were stained with haematoxylin and eosin (H&E). Immunohistochemistry was performed with an affinity purified anti-collectrin antibody. B⁰AT1 was detected using a rabbit anti-mouse B⁰AT1 antibody.

Amino acid transporter activities. Uptake of ³H-Phe, ¹⁴C-Gln and ³²PO₄ was determined in BBMVs at 60 s. The final amino acid concentration was 1 mM. Measurements were done in parallel in the presence of either Na⁺ or K⁺.

Expression studies and uptake measurements in *Xenopus laevis* oocytes and kidney MDCK cells were performed as described^{22,23}.

Received 5 May; accepted 23 November 2006.

Published online 13 December 2006.

- Imai, Y. *et al.* Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436**, 112–116 (2005).
- Crackower, M. A. *et al.* Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature* **417**, 822–828 (2002).
- Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nature Med.* **11**, 875–879 (2005).
- Zhang, H. *et al.* Collectrin, a collecting duct-specific transmembrane glycoprotein, is a novel homolog of ACE2 and is developmentally regulated in embryonic kidneys. *J. Biol. Chem.* **276**, 17132–17139 (2001).
- Fukui, K. *et al.* The HNF-1 target Collectrin controls insulin exocytosis by SNARE complex formation. *Cell Metabol.* **2**, 373–384 (2005).
- Akpınar, P., Kuwajima, S., Krutzfeldt, J. & Stoffel, M. Tmem27: A cleaved and shed plasma membrane protein that stimulates pancreatic β cell proliferation. *Cell Metabol.* **2**, 385–397 (2005).
- Schwenk, F., Baron, U. & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of *loxP*-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).
- Bergeron, A., Jorquera, R. & Tanguay, R. M. Hereditary tyrosinemia: an endoplasmic reticulum stress disorder? *Med. Sci.* **10**, 976–980 (2003).
- Endo, F. *et al.* Animal models reveal pathophysiologies of tyrosinemias. *J. Nutr.* **3**, 2063S–2067S (2003).
- Verrey, F. *et al.* Novel renal amino acid transporters. *Annu. Rev. Physiol.* **67**, 557–572 (2005).
- Kenny, A. J. & Maroux, S. Topology of microvillar membrane hydrolases of kidney and intestine. *Physiol. Rev.* **62**, 91–128 (1982).
- Romeo, E. *et al.* Luminal kidney and intestine SLC6 amino acid transporters of BOAT-cluster and their tissue distribution in *Mus musculus*. *Am. J. Physiol. Renal Physiol.* **290**, F376–F383 (2006).
- Broer, A. *et al.* Molecular cloning of mouse amino acid transport system B0, a neutral amino acid transporter related to Hartnup disorder. *J. Biol. Chem.* **279**, 24467–24476 (2004).
- Kleta, R. *et al.* Mutations in SLC6A19, encoding BOAT1, cause Hartnup disorder. *Nature Genet.* **6**, 999–1002 (2004).
- Seow, H. F. *et al.* Hartnup disorder is caused by mutations in the gene encoding the neutral amino acid transporter SLC6A19. *Nature Genet.* **6**, 1003–1007 (2004).
- Camargo, S. M., Makrides, V., Virkki, L. V., Forster, I. C. & Verrey, F. Steady-state kinetic characterization of the mouse BoAT1 sodium-dependent neutral amino acid transporter. *Eur. J. Physiol.* **451**, 338–348 (2005).
- Kowalczyk, S. *et al.* Molecular cloning of the mouse IMINO system: an Na⁺- and Cl⁻-dependent proline transporter. *Biochem. J. Mar.* **15**, 417–422 (2005).
- Peghini, P., Janzen, J. & Stoffel, W. Glutamate transporter EAAC-1-deficient mice develop dicarboxylic aminoaciduria and behavioral abnormalities but no neurodegeneration. *EMBO J.* **16**, 3822–3832 (1997).
- Feliubadaló, L. *et al.* Slc7a9-deficient mice develop cystinuria non-I and cystine urolithiasis. *Hum. Mol. Genet.* **12**, 2097–2108 (2003).
- Wagner, C. A., Lang, F. & Broer, S. Function and structure of heterodimeric amino acid transporters. *Am. J. Physiol. Cell Physiol.* **81**, C1077–C1093 (2001).
- Malathi, P., Preiser, H., Fairclough, P., Mallett, P. & Crane, R. K. A rapid method for the isolation of kidney brush border membranes. *Biochim. Biophys. Acta* **554**, 259–263 (1979).
- Ristic, Z. *et al.* Neutral amino acid transport mediated by ortholog of imino acid transporter SIT1/SLC6A20 in opossum kidney cells. *Am. J. Physiol. Renal Physiol.* **290**, F880–F887 (2006).
- Bauch, C., Forster, N., Löffing-Cueni, D., Summa, V. & Verrey, F. Functional cooperation of epithelial heteromeric amino acid transporters expressed in Madin–Darby canine kidney cells. *J. Biol. Chem.* **278**, 1316–1322 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. J. Crackower, A. Leibbrandt, Y. Imai, K. Kuba and many others for their contributions. Supported by grants from The National Bank of Austria, The Austrian Ministry of Science and Education, IMBA, an EU Marie Curie Excellence grant, and EUGeneHeart to J.M.P., from the Swiss National Science Foundation to F.V. and EUGINDAT to F.V. and C.A.W.

Author Contributions U.D., R.S., C.R., C.B., G.S., A.R., S.A., J.A.P., D.S., S.M.R.C., V.M., T.R. and C.A.W. performed experiments. U.D., F.V., C.A.W. and J.M.P. planned experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M.P. (josef.penninger@imba.oeaw.ac.at) or C.A.W. (wagnerca@access.unizh.ch).

LETTERS

Botulinum neurotoxin B recognizes its protein receptor with high affinity and specificity

Rongsheng Jin^{1,2}, Andreas Rummel³, Thomas Binz⁴ & Axel T. Brunger^{1,2}

Botulinum neurotoxins (BoNTs) are produced by *Clostridium botulinum* and cause the neuromuscular syndrome of botulism. With a lethal dose of 1 ng kg^{-1} , they pose a biological hazard to humans and a serious potential bioweapon threat¹. BoNTs bind with high specificity at neuromuscular junctions and they impair exocytosis of synaptic vesicles containing acetylcholine through specific proteolysis of SNAREs (soluble *N*-ethylmaleimide-sensitive fusion protein attachment protein receptors), which constitute part of the synaptic vesicle fusion machinery^{2,3}. The molecular details of the toxin–cell recognition have been elusive. Here we report the structure of a BoNT in complex with its protein receptor: the receptor-binding domain of botulinum neurotoxin serotype B (BoNT/B) bound to the luminal domain of synaptotagmin II, determined at 2.15 Å resolution. On binding, a helix is induced in the luminal domain which binds to a saddle-shaped crevice on a distal tip of BoNT/B. This crevice is adjacent to the non-overlapping ganglioside-binding site of BoNT/B. Synaptotagmin II interacts with BoNT/B with nanomolar affinity, at both neutral and acidic endosomal pH. Biochemical and neuronal *ex vivo* studies of structure-based mutations indicate high specificity and affinity of the interaction, and high selectivity of BoNT/B among synaptotagmin I and II isoforms. Synergistic binding of both synaptotagmin and ganglioside imposes geometric restrictions on the initiation of BoNT/B translocation after endocytosis. Our results provide the basis for the rational development of preventive vaccines or inhibitors against these neurotoxins.

BoNTs are initially synthesized as a single-chain precursor protein. After cleavage by clostridial or tissue proteases, the active two-chain neurotoxin is fully capable of performing neurospecific binding, receptor-mediated endocytosis, endosomal translocation and target proteolysis⁴. Seven structurally and functionally related BoNTs exist, termed serotypes A to G, which are also related to tetanus neurotoxin (TeNT). The highly specific and strong binding of BoNTs with their cell-surface targets involves protein and ganglioside receptors that specifically localize to the neuronal plasma membrane⁵. At present, the only protein receptors to have been identified are synaptotagmin I (Syt-I) and synaptotagmin II (Syt-II) for both BoNT/B and BoNT/G, and synaptic vesicle protein SV2 (isoforms A, B and C) for BoNT/A^{6–10}. Synaptotagmins are a family of transmembrane proteins that trigger Ca^{2+} -dependent neurotransmitter release. Syt-I and Syt-II are essential for synaptic transmission in neuromuscular junctions¹¹. Both BoNT/B and BoNT/G exploit the luminal domains of Syt-I and Syt-II when they are temporarily exposed on the cell surface as a result of exocytosis, and the toxin–receptor complexes are then taken up by neurons through the recycling of secretory vesicles. Although ganglioside-binding sites have been identified for some BoNTs^{12–14}, the protein-receptor-binding sites have been unknown up to now.

The carboxy-terminal domain of the heavy chain of BoNT/B (referred to as H_{CB}) is responsible for specific binding with the luminal domains of Syt-I and Syt-II (ref. 6). To facilitate crystallization, we designed a fusion protein between H_{CB} (residues 858–1291) and Syt-II (8–61) (referred to as H_{CB} –Syt-II; Supplementary Fig. 1a). We validated the biological relevance of this strategy by a series of biochemical and cell-based studies. First, the fusion protein (H_{CB} –Syt-II) is monodisperse and monomeric in solution, thus ruling out the existence of a *trans*-complex that might complicate crystallization (Supplementary Fig. 1b). Second, a fusion construct between Syt-II (8–61) and full-length BoNT/B no longer interacts with free Syt-II (1–61) and its toxicity is decreased more than tenfold, suggesting that the Syt-II-binding site on BoNT/B is occupied in the fusion protein (Supplementary Fig. 1c, d). Third, structure-based mutagenesis studies based on H_{CB} or full-length BoNT/B confirmed the importance of interacting residues for binding affinity and toxicity (described below). Taken together, these data show that the covalently linked Syt-II forms a *cis*-complex with BoNT/B that faithfully represents the binary complex.

The structure of the complex between BoNT/B and Syt-II was determined to a minimum Bragg spacing of $d_{\text{min}} = 2.15 \text{ Å}$ by molecular replacement with the structure of H_{CB} (Protein Data Bank code 1Z0H) as the search model¹⁵ (Supplementary Table 1). A difference electron-density map allowed unambiguous assignment of Syt-II residues 44–60, which are structured in the complex (Fig. 1a). This is consistent with the previous finding that residues 40–60 of Syt-II comprise a minimal BoNT/B-binding domain⁹. Residues 8–43 of Syt-II and the linker were not visible, probably as a result of conformational variability. The C terminus of H_{CB} is relatively close to the observed residue Glu 44 of Syt-II (about 41 Å; Fig. 1b), with the linker and Syt-II residues 8–43 providing a sufficient number of residues to connect the termini. Many possible paths between the termini are unobstructed by crystal packing, including those in a second structure we obtained in a different crystal form with a different packing arrangement at 3.2 Å resolution (not shown). We conclude that the linker does not affect the structure of the complex.

The luminal domain of Syt-II is unstructured in solution (Fig. 1d). On binding to H_{CB} , residues Glu 44 to Lys 60 become structured, with residues Phe 47 to Ile 58 forming an α -helix (Fig. 1a, b). Syt-II binds at the distal tip of the C-terminal domain of H_{CB} , in a saddle-shaped crevice on the surface formed by eight β -strands and four intervening loops. The strand that forms the base of the crevice is parallel to the Syt-II helix (Fig. 1b). The extensive intermolecular interface buries a solvent-accessible surface area of about $1,200 \text{ Å}^2$, involving mostly hydrophobic residues, with two pronounced pockets on the surface of H_{CB} (Fig. 2, and Supplementary Fig. 2). In addition, H_{CB} residues Lys 1113, Ser 1116 and Lys 1192 form salt

¹Howard Hughes Medical Institute, ²Departments of Molecular and Cellular Physiology, Neurology and Neurological Science, Structural Biology, and Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford, California 94305, USA. ³Institut für Toxikologie, ⁴Institut für Biochemie, Medizinische Hochschule Hannover, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany.

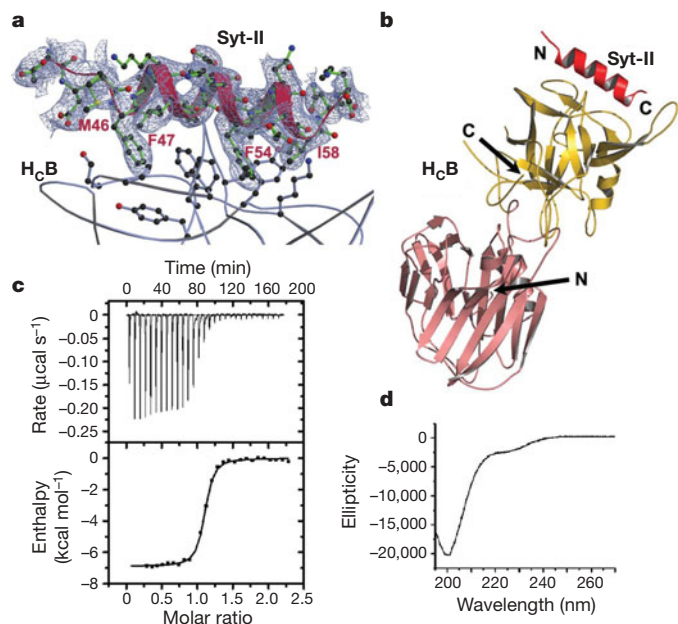


Figure 1 | Structure of the H_CB–Syt-II complex. **a**, σ_A -weighted $F_o - F_c$ electron-density map (contoured at 1.5σ) around Syt-II, overlaid with the final refined model (red and green, Syt-II; grey, H_CB). This map is model-bias free because it is calculated from the phases of the atomic model before the inclusion of the Syt-II peptide (using a lower-resolution diffraction data set to 2.6 Å). At this stage, the R_{work} and R_{free} values were 26.1% and 31.6%, respectively. **b**, Structure of the complex between H_CB (pink and gold) and Syt-II (red). **c**, Titration of GST–Syt-II (1–61) into H_CB by ITC at 20 °C and pH 7.5. $N = 1.1$; $K_d = 2.64 \times 10^7 \text{ M}^{-1}$; $\Delta H = -6,890 \text{ cal mol}^{-1}$; $\Delta S = 10.4 \text{ cal mol}^{-1} \text{ K}^{-1}$. **d**, Far-ultraviolet circular dichroism spectra of the luminal domain of Syt-II (1–61).

bridges with residue Glu 57 of Syt-II. The structure of Syt-II-bound H_CB is very similar to that of the apo state with an overall root-mean-square deviation of 0.73 Å (ref. 15), with the notable exception of H_CB residues Glu 1191, Tyr 1183 and Phe 1204, which undergo significant rotamer changes on binding.

We measured the thermodynamics of the assembly of the H_CB–Syt-II complex in solution by using isothermal titration calorimetry (ITC). The tight binding of H_CB to Syt-II (average $K_d \approx 34 \text{ nM}$) is stoichiometric (about 1:1), endothermic (average $\Delta H \approx -7.4 \text{ kcal mol}^{-1}$; $1 \text{ cal} \approx 4.18 \text{ J}$) and entropy driven (average $\Delta S \approx 9.0 \text{ cal mol}^{-1} \text{ K}^{-1}$) (Fig. 1c, and Supplementary Table 2). The heat capacity (ΔC_p)

for the H_CB–Syt-II interaction is about $-326 \text{ cal mol}^{-1} \text{ K}^{-1}$, which is consistent with a protein–protein interaction driven by the hydrophobic effect^{16,17}. ITC titration at pH 5.7, mimicking the acidic endosomal environment, produced no change on assembly thermodynamics, indicating that the pH change associated with toxin internalization is unlikely to affect the binding of BoNT/B to its protein receptor (Supplementary Fig. 3b and Supplementary Table 2).

To study the specificity of the BoNT/B–Syt-II interaction, we performed structure-based mutagenesis studies both *in vitro* and in a neuronal assay. Point mutations were introduced in H_CB or Syt-II at or around the toxin–receptor interface. None of the H_CB single-site mutations caused an effect on protein folding and stability as verified by far-ultraviolet circular dichroism spectroscopy and thermal denaturation experiments. Most of the mutations in H_CB (V1118D, K1192E, F1194A, A1196K and F1204A) and in Syt-II (F47A, F54A, F55A and E57K) greatly decrease binding affinity (Fig. 3a, d), which is consistent with the significance of these residues for the H_CB–Syt-II interaction (Fig. 2). The relatively minor contribution of Met 46 and Leu 50 of Syt-II to the affinity is consistent with their locations at the periphery of the interface. The mutation of a residue located on the opposite side of the toxin–receptor interface, Syt-II-N59H, behaves similarly to the wild type. H_CB-S1199Y shows enhanced binding to Syt-II, perhaps as a result of a π -stacking interaction between Tyr 1199 of H_CB and Phe 47 of Syt-II.

We further examined the effects of the mutants in BoNT/B on physiological function—that is, toxicity—by using recombinant full-length single-chain BoNT/B applied at the mouse phrenic nerve^{6,18} (Fig. 3b). The decrease in toxicity is marked (up to about 1,000-fold) for all mutations that reduce the interaction as shown by the pull-down experiments. The effect is comparable to or larger than that observed for mutations in the ganglioside-binding site¹⁴. It is remarkable that single-site mutants at the BoNT/B–Syt-II interface significantly attenuate the BoNT/B toxicity, suggesting that this protein–protein interface can be a target for the development of therapeutics to block the action of BoNT/B.

Syt-II alone is sufficient to mediate the binding and entry of BoNT/B into nerve cells, whereas Syt-I acts in a ganglioside-dependent manner⁹, even though Syt-I and Syt-II are very similar in both primary sequence and function¹⁹. Furthermore, the dissociation constant (K_d) between H_CB and the luminal domain of Syt-I is at least two orders of magnitude larger than that between H_CB and the luminal domain of Syt-II (ITC data not shown). Residues of Syt that interact with H_CB are conserved in both Syt-I and Syt-II isoforms, except for Met 46, Phe 55 and Ile 58 (Fig. 3c). Because Met 46 of Syt-II contributes little to BoNT/B binding, we introduced Syt-II-like side chains

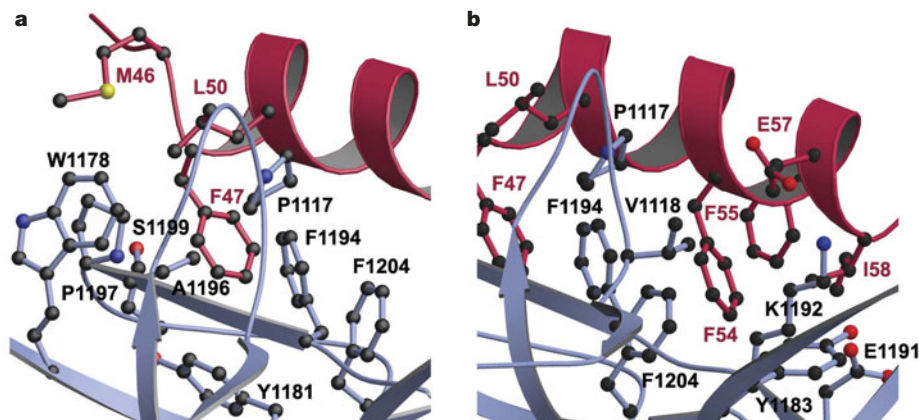


Figure 2 | The H_CB–Syt-II complex is stabilized by extensive intermolecular interactions involving two pronounced pockets on the H_CB surface. **a**, Pocket formed by H_CB residues Pro 1117, Trp 1178, Tyr 1181, Phe 1194, Ala 1196, Pro 1197 and Phe 1204, to which Syt-II residues Met 46, Phe 47 and Leu 50 bind. **b**, Pocket formed by H_CB residues Lys 1113,

Ser 1116, Pro 1117, Val 1118, Tyr 1183, Glu 1191, Lys 1192, Phe 1194 and Phe 1204. Four Syt-II residues, Phe 54, Phe 55, Glu 57 and Ile 58 interact with this pocket. The complex of Syt-II (red) with H_CB (grey) is oriented similarly to that shown in Fig. 1a. Key residues are in ball-and-stick representation.

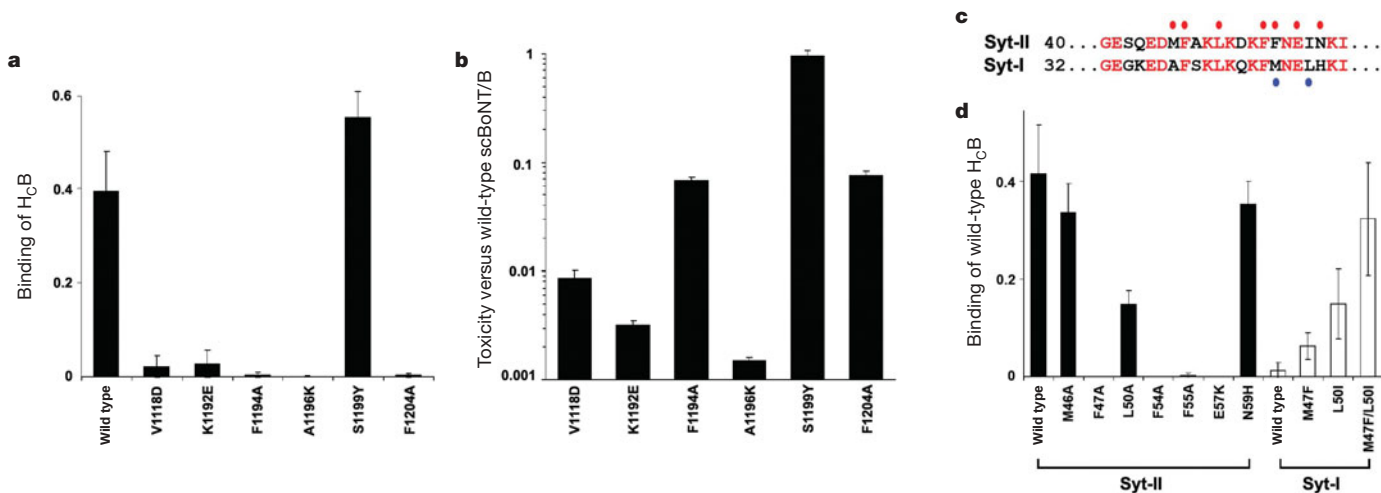


Figure 3 | Site-directed mutagenesis analysis of the toxin-receptor binding site in BoNT/B and in the luminal domains of Syt-I and Syt-II. **a**, Pull-down assay using GST-Syt-II wild type as bait in the presence of mixed ganglioside/Triton X-100 micelles shows drastically reduced binding of various $H_C B$ single-site mutants. **b**, The corresponding single-site mutations in recombinant full-length single-chain BoNT/B confirm the drastic decrease in toxicity employing the mouse phrenic nerve as an *ex vivo* system. Results are plotted on a logarithmic scale. **c**, Sequence alignment of the

membrane-proximal 22 amino acids of the luminal domain of rat Syt-I and Syt-II. The non-conserved residues are shown in black. Mutated residues of Syt-I and Syt-II are indicated with blue dots and red dots, respectively.

d, Pull-down assay with $H_C B$ wild-type and mutants of GST-Syt-II (1–61) (filled columns) and GST-Syt-I (1–53) (open columns) as bait in the presence of mixed ganglioside/Triton X-100 micelles shows the crucial role of Phe 47, Phe 54, Phe 55 and Glu 57 in Syt-II. Values in **a**, **b** and **d** are means $\pm$ s.d.

into Syt-I at the equivalent positions of residues 55 and 58. The double mutation (M47F/L50I) converts Syt-I into a Syt-II-like high-affinity receptor (Fig. 3d). The three BoNT/B residues (Glu 1191, Tyr 1183 and Phe 1204) that undergo rotamer changes on receptor binding are part of the pocket that interacts with these two Syt residues (Fig. 2b), indicating that this pocket of BoNT/B might have the key function in determining the protein-receptor-binding specificity.

Our results have implications for the function of BoNT/G, the closest homologue to BoNT/B, especially considering that the Syt-II-binding site is mostly conserved in BoNT/B and BoNT/G (Supplementary Fig. 4). BoNT/G also binds to the membrane-proximal region of Syt-I and Syt-II (ref. 6). Mutations of some of the BoNT/G residues that are equivalent to the Syt-interacting residues on BoNT/B significantly decrease the binding affinities between Syt-I and Syt-II and BoNT/G (ref. 20).

The two-receptor hypothesis for BoNTs was first proposed 20 years ago²¹, although the relationship between these two receptors has been unclear. We now find that Syt-II and ganglioside occupy

two adjacent but non-overlapping binding sites^{12–14} (Fig. 4). The glucose of sialyllactose and the C terminus of the Syt-II luminal domain both point towards the membrane (Fig. 4). Sialyllactose (up to 12 μ M) did not show a measurable effect on the binding between $H_C B$ and the luminal domain of Syt-I as examined by ITC (data not shown). Deletion of the transmembrane domain of Syt-I abolishes ganglioside-dependent binding²². Thus, gangliosides do not enhance BoNT/B–Syt-I binding in an allosteric manner. However, the two receptors act synergistically in the context of the membrane because the dissociation constant between $H_C B$ and the luminal domain of Syt in solution is more than 100-fold larger than that measured between BoNT/B and the full-length Syt (including the transmembrane region) in the presence of gangliosides and micelles⁷. Nevertheless, the interactions may be different for other members of the clostridial neurotoxin family because two carbohydrate-binding sites in the C-terminal part of the H_C fragment of TeNT are required for its function²³. These different mechanisms of cell recognition probably cause their different targeting and sorting in peripheral neurons.

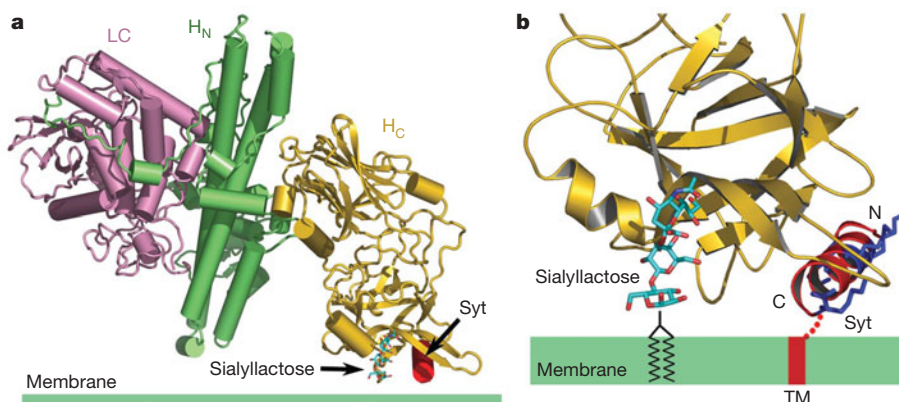


Figure 4 | Simultaneous binding with membrane-anchored Syt-II and ganglioside imposes geometric restrictions on how BoNT/B binds to the membrane surface. **a**, Proposed binding mode of BoNT/B on the membrane surface. The structure of a sialyllactose-bound BoNT/B (PDB code 1F31) was superimposed on the complex of $H_C B$ –Syt-II by using the coordinates of

the H_C fragment for the alignment. The light chain (LC), the N-terminal part of the heavy chain (H_N), and the C-terminal domain of the heavy chain (H_C) are shown in pink, green and gold, respectively. **b**, A close-up view of the proposed interface between BoNT/B and membrane. Four lysine residues that are conserved in Syt-I and Syt-II are coloured blue.

For BoNTs, a proper orientation on the membrane surface is important for efficient endocytosis and subsequent translocation of the light chain to the cytosol^{24,25}. The simultaneous attachment of Syt and ganglioside imposes geometric restrictions on the position of BoNT/B with respect to the membrane surface (Fig. 4). The two mostly negatively charged molecular surfaces, which remain negatively charged even at luminal pH^{26,27}, further restrict the orientation of BoNT/B on the membrane surface (Supplementary Fig. 5a, b). In addition, four solvent-exposed lysine residues (Lys 49, Lys 51, Lys 53 and Lys 60) are conserved in both Syt-I and Syt-II (Fig. 3c) and might interact with the phospholipid headgroup region. BoNT/B probably adopts a similar binding mode on the membrane after the initiation of endosomal translocation, because the binding affinity between H₂C_B and the isolated Syt-II is independent of pH.

Given the safety concern about the currently used equine antitoxin therapy and the scarcity of the toxoid¹, effective binding inhibitors or vaccines are urgently needed against botulism. Our results are a step towards the development of safer vaccines, for example by using recombinant BoNT/B with structure-based mutations in the critical protein-receptor-binding site. The structure of the complex could also be used as a starting point for pharmacological inhibitor design.

METHODS

Detailed methods are provided in the Supplementary Information. In brief, the H₂C_B-Syt-II fusion protein was obtained from the H₂C_B fragment of BoNT/B (H₂C_B, residues 858–1291) and the luminal domain of rat Syt-II (residues 8–61), connected by a 15-residue linker (PPTPGSAWSHPQFEK) including a Strep-tag (Supplementary Fig. 1a). Crystals were grown at 20 °C by vapour diffusion in two conditions: first, 13% polyethylene glycol (PEG) 6,000 and 0.1 M HEPES pH 7.0, and second, 0.8 M sodium citrate pH 6.5. The PEG condition yielded higher-quality crystals that diffracted to $d_{\min} = 2.15$ Å. This crystal form adopts space group $P2_12_12_1$, with unit cell dimensions $a = 55.8$ Å, $b = 97.5$ Å, $c = 113.6$ Å, and one H₂C_B-Syt-II complex in the asymmetric unit. The structure of the H₂C_B-Syt-II complex was solved by molecular replacement with H₂C_B as the search model (PDB code 1Z0H) using Phaser²⁸. The structure was built with the use of COOT²⁹ and refined to final $R_{\text{work}} = 19.9\%$ and $R_{\text{free}} = 24.5\%$ at 2.15 Å resolution with CNS³⁰. The interactions between wild-type Syt-II and H₂C_B mutants, Syt-II mutants and wild-type H₂C_B, or Syt-I mutants and wild-type H₂C_B, were examined by a glutathione S-transferase (GST) pull-down assay. ITC experiments were conducted between H₂C_B and GST-Syt-II or Syt-II alone. Titrations were performed at different temperatures, protein concentrations and pH values. All H₂C_B mutations were further incorporated into full-length BoNT/B and the toxicities were determined by a mouse phrenic nerve toxicity assay¹⁸. The measured paralytic half-times were converted to toxicities against the corresponding wild-type BoNT/B.

Received 20 August; accepted 30 October 2006.

Published online 13 December 2006.

1. Arnon, S. S. *et al.* Botulinum toxin as a biological weapon: medical and public health management. *J. Am. Med. Assoc.* **285**, 1059–1070 (2001).
2. Schiavo, G. *et al.* Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359**, 832–835 (1992).
3. Chen, Y. A., Scales, S. J., Patel, S. M., Doung, Y. C. & Scheller, R. H. SNARE complex formation is triggered by Ca^{2+} and drives membrane fusion. *Cell* **97**, 165–174 (1999).
4. Montecucco, C. & Schiavo, G. Structure and function of tetanus and botulinum neurotoxins. *Q. Rev. Biophys.* **28**, 423–472 (1995).
5. Montecucco, C., Rossetto, O. & Schiavo, G. Presynaptic receptor arrays for clostridial neurotoxins. *Trends Microbiol.* **12**, 442–446 (2004).
6. Rummel, A., Karnath, T., Henke, T., Bigalke, H. & Binz, T. Synaptotagmins I and II act as nerve cell receptors for botulinum neurotoxin G. *J. Biol. Chem.* **279**, 30865–30870 (2004).
7. Nishiki, T. *et al.* The high-affinity binding of *Clostridium botulinum* type B neurotoxin to synaptotagmin II associated with gangliosides GT1b/GD1a. *FEBS Lett.* **378**, 253–257 (1996).
8. Dong, M. *et al.* SV2 is the protein receptor for botulinum neurotoxin A. *Science* **312**, 592–596 (2006).
9. Dong, M. *et al.* Synaptotagmins I and II mediate entry of botulinum neurotoxin B into cells. *J. Cell Biol.* **162**, 1293–1303 (2003).

10. Mahrhold, S., Rummel, A., Bigalke, H., Davletov, B. & Binz, T. The synaptic vesicle protein 2C mediates the uptake of botulinum neurotoxin A into phrenic nerves. *FEBS Lett.* **580**, 2011–2014 (2006).
11. Pang, Z. P. *et al.* Synaptotagmin-2 is essential for survival and contributes to Ca^{2+} -triggering of neurotransmitter release in central and neuromuscular synapses. *J. Neurosci.* (in the press).
12. Eswaramoorthy, S., Kumaran, D. & Swaminathan, S. Crystallographic evidence for doxorubicin binding to the receptor-binding site in *Clostridium botulinum* neurotoxin B. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1743–1746 (2001).
13. Swaminathan, S. & Eswaramoorthy, S. Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B. *Nature Struct. Biol.* **7**, 693–699 (2000).
14. Rummel, A., Mahrhold, S., Bigalke, H. & Binz, T. The H₂C_B-domain of botulinum neurotoxins A and B exhibits a singular ganglioside binding site displaying serotype specific carbohydrate interaction. *Mol. Microbiol.* **51**, 631–643 (2004).
15. Jayaraman, S., Eswaramoorthy, S., Ahmed, S. A., Smith, L. A. & Swaminathan, S. N-terminal helix reorients in recombinant C-fragment of *Clostridium botulinum* type B. *Biochem. Biophys. Res. Commun.* **330**, 97–103 (2005).
16. Perozzo, R., Folkers, G. & Scapozza, L. Thermodynamics of protein–ligand interactions: history, presence, and future aspects. *J. Recept. Signal Transduct. Res.* **24**, 1–52 (2004).
17. Stites, W. E. Protein–protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* **97**, 1233–1250 (1997).
18. Habermann, E., Dreyer, F. & Bigalke, H. Tetanus toxin blocks the neuromuscular transmission *in vitro* like botulinum A toxin. *Naunyn Schmiedeberg's Arch. Pharmacol.* **311**, 33–40 (1980).
19. Südhof, T. C. Synaptotagmins: why so many? *J. Biol. Chem.* **277**, 7629–7632 (2002).
20. Rummel, A. *et al.* Identification of the protein receptor binding site of botulinum neurotoxins B and G proves the double receptor concept. *Proc. Natl Acad. Sci. USA* (in the press).
21. Montecucco, C. How do tetanus and botulinum neurotoxins bind to neuronal membranes? *Trends Biochem. Sci.* **11**, 314–317 (1986).
22. Kozaki, S., Kamata, Y., Watarai, S., Nishiki, T. & Mochida, S. Ganglioside GT1b as a complementary receptor component for *Clostridium botulinum* neurotoxins. *Microb. Pathog.* **25**, 91–99 (1998).
23. Rummel, A., Bade, S., Alves, J., Bigalke, H. & Binz, T. Two carbohydrate binding sites in the H₂C_B-domain of tetanus neurotoxin are required for toxicity. *J. Mol. Biol.* **326**, 835–847 (2003).
24. Koriazova, L. K. & Montal, M. Translocation of botulinum neurotoxin light chain protease through the heavy chain channel. *Nature Struct. Biol.* **10**, 13–18 (2003).
25. Hoch, D. H. *et al.* Channels formed by botulinum, tetanus, and diphtheria toxins in planar lipid bilayers: relevance to translocation of proteins across membranes. *Proc. Natl Acad. Sci. USA* **82**, 1692–1696 (1985).
26. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
27. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins* **61**, 704–721 (2005).
28. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
29. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
30. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Adams and J. Zhou for critical reading of the manuscript, and the staff of beamlines 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) and beam line 8.2.2 at the Advanced Light Source (ALS) for help during data collection; H. Bigalke for providing the mouse phrenic nerve facility; and T. Henke and C. Knorr for technical assistance. The SSRL is a national user facility operated by Stanford University on behalf of the US Department of Energy (Office of Basic Energy Sciences). The SSRL Structural Molecular Biology Program is supported by the Department of Energy (Office of Biological and Environmental Research), and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program), and the National Institute of General Medical Sciences. The ALS is supported by the Office of Energy Research (Office of Basic Energy Sciences, Material Sciences Division) of the US Department of Energy at Lawrence Berkeley National Laboratory. Support by the Department of Defense and Defense Threat Reduction Agency (to A.T.B.) and by a Deutsche Forschungsgemeinschaft grant (to T.B.) is acknowledged.

Author Information The atomic coordinates and structure factors of the H₂C_B-Syt-II complex are deposited in the Protein Data Bank under accession code 2NM1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.T.B. (brunger@stanford.edu).

LETTERS

Structural basis of cell surface receptor recognition by botulinum neurotoxin B

Qing Chai^{1*}, Joseph W. Arndt^{1*}, Min Dong^{2*}, William H. Tepp³, Eric A. Johnson³, Edwin R. Chapman² & Raymond C. Stevens¹

Botulinum neurotoxins (BoNTs) are potent bacterial toxins that cause paralysis at femtomolar concentrations¹ by blocking neurotransmitter release. A 'double receptor' model has been proposed in which BoNTs recognize nerve terminals via interactions with both gangliosides and protein receptors that mediate their entry². Of seven BoNTs (subtypes A–G), the putative receptors for BoNT/A^{3,4}, BoNT/B^{5,6} and BoNT/G⁷ have been identified, but the molecular details that govern recognition remain undefined. Here we report the crystal structure of full-length BoNT/B in complex with the synaptotagmin II (Syt-II) recognition domain at 2.6 Å resolution. The structure of the complex reveals that Syt-II forms a short helix that binds to a hydrophobic groove within the binding domain of BoNT/B. In addition, mutagenesis of amino acid residues within this interface on Syt-II affects binding of BoNT/B. Structural and sequence analysis reveals that this hydrophobic groove is conserved in the BoNT/G and BoNT/B subtypes, but varies in other clostridial neurotoxins. Furthermore, molecular docking studies using the ganglioside G_{T1b} indicate that its binding site is more extensive than previously proposed and might form contacts with both BoNT/B and synaptotagmin. The results provide structural insights into how BoNTs recognize protein receptors and reveal a promising target for blocking toxin–receptor recognition.

Each BoNT is composed of a light chain of ~50 kDa that acts as a metalloprotease, and a heavy chain of ~100 kDa, which includes an amino-terminal domain (H_N) that mediates the translocation of the light chain from endosomes into cytosol, and a carboxy-terminal domain (H_C) that mediates the binding to specific cell surface receptors. It has been established that BoNT/A uses three isoforms of synaptic vesicle 2 (SV2A, SV2B and SV2C)^{3,4}, and BoNT/B uses two homologues of synaptotagmin (Syt-I and Syt-II) as their protein receptors^{5,6,8}, respectively. Additionally, BoNT/G also binds Syt-I and Syt-II and may use them as receptors⁷. SV2 and Syt-I/Syt-II are synaptic vesicle membrane proteins. Their toxin-recognition domains are located in the vesicle lumen and are exposed to the extracellular milieu transiently after synaptic vesicles fuse with the plasma membrane.

We initiated our structural studies by co-crystallizing full-length BoNT/B with the peptide bearing the toxin recognition domain of Syt-II (residues 40–60)⁶ (Supplementary Fig. 1). The structure of the complex was determined at 2.6 Å resolution by molecular replacement (Fig. 1; Supplementary Table 1). The crystal structure reveals that Syt-II binds to a hydrophobic groove in the C-terminal trefoil subdomain (H_{CC}) of the H_C domain, primarily composed of loops 1115, 1185, 1190 and 1200; the binding site contains two pockets that are divided by a shallow ridge (Supplementary Fig. 2). The recog-

nition domain of Syt-II docks into the groove as a short helix, making extensive hydrophobic contacts with BoNT/B through 607.9 Å² of buried surface area at the interface upon formation of the complex (Fig. 2a).

Some notable structural features of this interface include the complementary hydrophobic and π -stacking contacts made by the side chains F47, L50, F54, F55 and I58 of Syt-II that bury into the binding groove of BoNT/B (Fig. 2b). Additionally, the charged side chains K51 and E57 of Syt-II are close to the oppositely charged residues E1202 and K1112 of BoNT/B, respectively, and probably form electrostatic interactions to further stabilize the complex (Supplementary Table 2). Lastly, hydrogen-bonding interactions are found between K53 and E57 of Syt-II with D1114 (main-chain carbonyl oxygen) and S1115 of BoNT/B, correspondingly. The overall structure of BoNT/B in the complex is very similar to the isolated holotoxin structure (PDB accession number 1EPW⁹) with a C α root-mean-square deviation of 1.05 Å. Interestingly, the recognition domain of Syt-II does not form an obvious secondary structure in the absence of BoNT/B, as we determined using circular dichroism spectroscopy (Supplementary Fig. 3a), suggesting that the helical conformation of Syt-II is induced upon BoNT/B binding. An induced-fit formation of a helix was also identified in SNAP-25

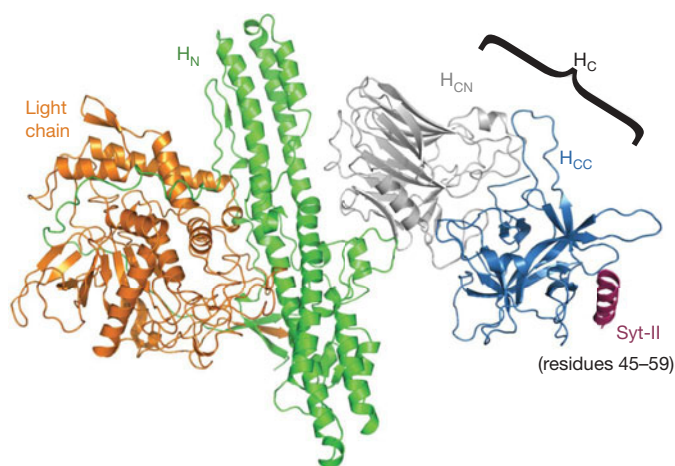


Figure 1 | The structure of BoNT/B–Syt-II recognition domain complex. Ribbon diagram of BoNT/B–Syt-II recognition domain complex, with subdomains of BoNT/B labelled as light chain (orange), H_N (green), H_{CN} (grey), and H_{CC} (blue). The Syt-II recognition domain (residues 45–59, magenta) forms an α -helix when complexed with BoNT/B.

¹Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA. ²Howard Hughes Medical Institute and Department of Physiology, 1300 University Avenue, and ³Food Research Institute, Department of Food Microbiology and Toxicology, 1925 Willow Drive, The University of Wisconsin, Madison, Wisconsin 53706, USA.

*These authors contributed equally to this work.

(synaptosome-associated protein 25 kDa) upon binding by the BoNT/A light chain in the crystal structure of the BoNT/A-light-chain–SNAP-25 complex¹⁰. The hydrophobic interactions between BoNT/B and Syt-II probably drive the helix formation that occurs in the Syt-II recognition domain to overcome the energy cost for folding.

Residues 40–60 of Syt-II were found to be the region governing binding⁶ but here we further restrict it to residues 47–60, as indicated by independent biochemical binding assays (Fig. 2c). This correlates well with the structural identification that only residues 45–59 of Syt-II interact with BoNT/B, with the two N-terminal residues having weak electron density. Both Syt-I and Syt-II bind with BoNT/B, yet their dissociation constants are 2.3 and 0.23 nM, respectively⁵. To investigate the binding affinity and establish which residues are important for binding, we performed targeted mutagenesis on Syt-II. Mutants F47A, F54A and F55A of Syt-II abolish binding with BoNT/B, while E52A does not affect binding (Fig. 2d). In addition, we did pairwise mutagenesis by exchanging the five residues within the toxin recognition domain between Syt-II and Syt-I (Supplementary Fig. 3b). Of the resulting Syt-II mutants, A48S, E52Q and F55M still bind BoNT/B in the absence of gangliosides (Fig. 2e). However, exchanging I58 and N59 of Syt-II to the corresponding residues of Syt-I (L50 and H51) created a mutant that bound to BoNT/B only in the presence of gangliosides; this Syt-II mutant behaves like Syt-I in the biochemical binding assay (Fig. 2e). Conversely, substitution of residues L50 and H51 of Syt-I with the corresponding residues of Syt-II (I58 and N59) resulted in a Syt-I mutant that binds to BoNT/B in the absence of gangliosides (Fig. 2f).

Like the structural data, these results suggest the side chains of F47, F54 and F55 are buried in the hydrophobic pockets of the BoNT/B binding groove, while residues A48 and E52 are not engaged in the toxin–receptor interface; the conservative substitution of F55M

maintains hydrophobic contacts without compromising binding of BoNT/B. When modelled on the structure of the BoNT/B–Syt-II complex, the double mutant (I58L,N59H) introduces a steric clash with E1190 of BoNT/B. The weakened interaction between the Syt-II mutant (I58L,N59H) and BoNT/B thus requires ‘assistance’ from gangliosides. We conclude that conservative residues F54, F55 and F47 are critical for binding; and residues K51, K53 and E57 further strengthen the interaction. Subtle disparities between L50 and H51 of Syt-I and I58 and N59 of Syt-II are likely to account for their different binding affinities, providing the structural basis that may explain the different sensitivities of neuronal subtypes to BoNT/B¹¹. It will be interesting to determine whether polymorphisms exist in the recognition domain of Syt-I/II in humans—alterations in this region might render some patients resistant to BoNT/B treatment.

We have presented crystallographic data describing the molecular details of toxin–receptor recognition of a clostridial neurotoxin. In addition to providing structural insights into receptor recognition by BoNT/B, this study provides a molecular basis for studying other potential neurotoxin–receptor interactions. First, a structural comparison of the Syt-II binding site of BoNT/B with two tetanus neurotoxin structures in complex with a tri-peptide (YEW) or a disialyllactose¹² reveals a common ligand-binding groove (Supplementary Fig. 4). Although the tri-peptide or disialyllactose binding site within the H_{CC} domain of tetanus neurotoxin is analogous in location to the Syt-II binding site of BoNT/B, the chemical identities of these binding sites are very different. The Syt-II binding site on BoNT/B is a generously opened groove, displaying marked hydrophobic patches (Fig. 2a), with a mostly neutral electrostatic potential surface (Supplementary Fig. 4a). In contrast, the tetanus neurotoxin binding site for YEW or disialyllactose is relatively small, revealing more hydrophilic and basic residues at its interface (Supplementary Fig. 4b).

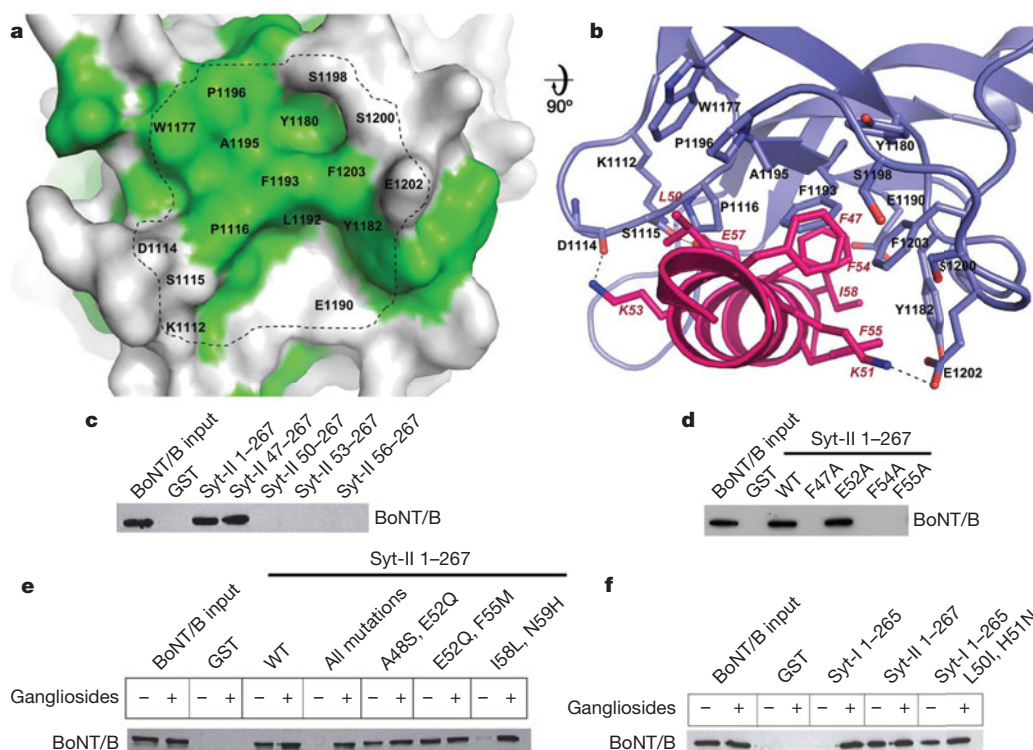


Figure 2 | Interaction between the H_{CC} domain of BoNT/B and the recognition domain of Syt-II. **a**, Binding interface on BoNT/B (circled in dashed line) reveals hydrophobic groove (green, hydrophobic surface). **b**, Close-up view of the binding interface of Syt-II (magenta) and BoNT/B (light blue) by rotating 90° around the x axis. **c**, Truncation mutants of Syt-II

were tested for BoNT/B binding activity. **d**, Point mutants of Syt-II were tested for BoNT/B binding activity. **e**, **f**, Various synaptotagmin mutants, in which residues within the recognition domain were interconverted between Syt-II and Syt-I, were tested for BoNT/B binding activity. WT, wild type; GST, glutathione S-transferase.

Second, we carried out sequence analyses of all BoNT/B subtypes, BoNT/G, and all BoNT/A subtypes, in an effort to address the potential interaction of these toxins with their individual receptors. We examined the hydrophobic residues that make key contacts with Syt-II in BoNT/B and found that most residues are conserved in BoNT/G (Supplementary Fig. 5a). This correlates with the finding that both BoNT/B and BoNT/G interact with a similar region of synaptotagmin: 10–14 luminal residues juxtaposed to the transmembrane domain⁷. Likewise, all subtypes of BoNT/B may use Syt-I/Syt-II as receptors, because the residues that make key contacts are highly conserved within BoNT/B subtypes (Supplementary Fig. 5a). Furthermore, we identified the cognate residues of the BoNT/A subtypes that correspond to the Syt-II binding site of BoNT/B and mapped them onto the structure of BoNT/A1 (3BTA¹³). These residues cluster together, also displaying notable hydrophobicity (Supplementary Fig. 5b). It will be interesting to determine whether hydrophobic interactions play a major role in BoNT/A–SV2 recognition. Lastly, we plotted the sequence conservation among all clostridial neurotoxins onto the surface of the H_{CC} domain of BoNT/B. The high surface residue variability at the corresponding synaptotagmin binding site (Supplementary Fig. 5c) implicates structural divergence, which we would expect to contribute to the diversity of receptor recognition.

As the other component of the ‘double receptor’ model, gangliosides are essential for the binding and functional entry of BoNT/B into cells^{14–18}. We further investigated the role of gangliosides using PC12 cells that stably express Syt-II⁶. After depleting endogenous ganglioside by PPMP (D,L-threo-1-phenyl-2-hexadecanoylamino-3-morpholino-propanol-HCl) treatment¹⁹, the level of binding and

internalization of BoNT/B is dramatically decreased. Loading with exogenous gangliosides restores BoNT/B binding and internalization, as evaluated by immunostaining of vesicle-associated membrane protein 2 (VAMP-2, also known as synaptobrevin 2), with loss of signal as a measure of proteolytic cleavage (Fig. 3a, b). Of four different gangliosides tested, G_{T1b} shows the highest level of facilitating BoNT/B binding and entry, followed by G_{D1a}, then G_{D1b}, and G_{M1} has the least effect. A similar pattern of ganglioside dependence is observed using recombinant Syt-II and Syt-I fusion proteins in BoNT/B pull-down assays (Fig. 3c). These findings are consistent with previous reports that G_{T1b} and G_{D1a} support binding of BoNT/B to Syt-II-containing liposomes and cell surface membranes, while G_{D1b} and G_{M1} failed to do so^{5,8,16}. Interestingly, there is a terminal sialic acid (Sia5) group present only in G_{T1b} and G_{D1a}, but not in G_{D1b} and G_{M1} (Supplementary Fig. 6) that might be important in ganglioside-supported BoNT/B binding with synaptotagmin.

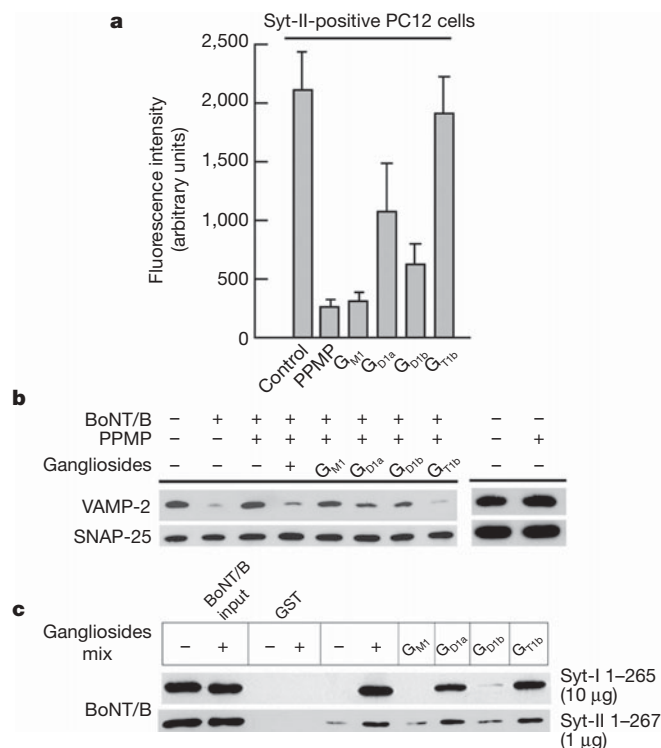


Figure 3 | Ganglioside-supported BoNT/B binding with synaptotagmin. **a**, PPMP treatment, which inhibits ganglioside synthesis¹⁹, strongly reduced binding of BoNT/B to Syt-II-positive PC12 cells. Binding of BoNT/B was restored with exogenous G_{T1b} as measured by immunostaining. Other gangliosides (G_{D1a} and G_{D1b}) exhibit moderate abilities to restore binding of BoNT/B, while G_{M1} had no effect. Error bars represent the standard deviation ($n = 20$). **b**, Entry of BoNT/B into cells was detected by western blot analysis using an antibody against VAMP-2. SNAP-25 is shown as a control. **c**, Gangliosides facilitate binding of BoNT/B with recombinant Syt-I and Syt-II as shown by pull-down assays.

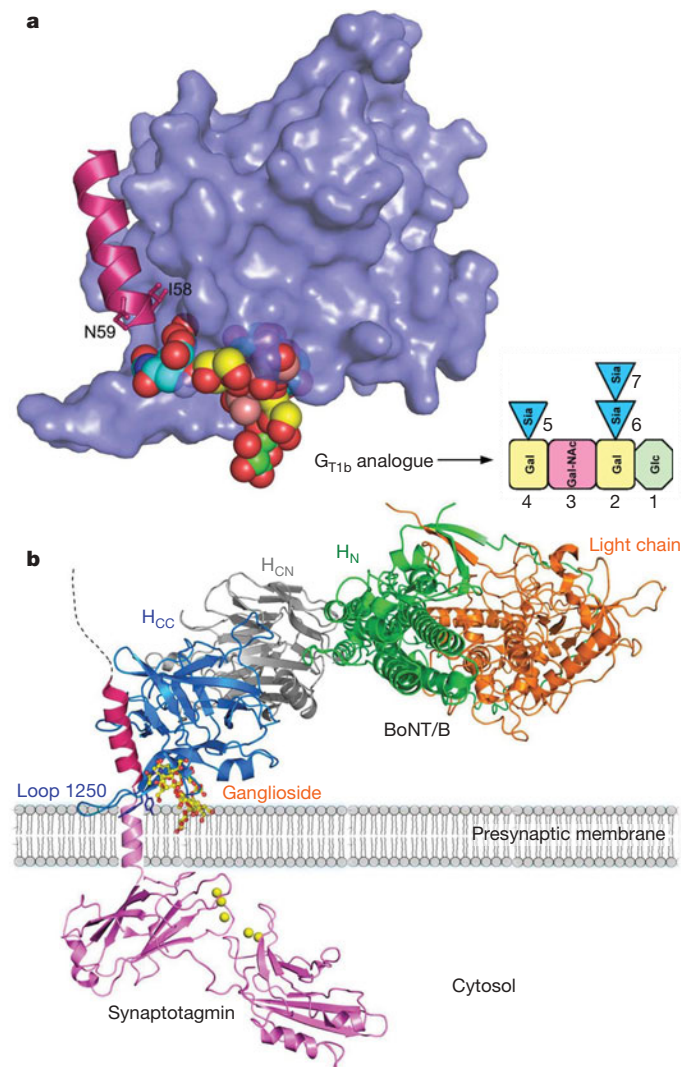


Figure 4 | Model of BoNT/B binding to its receptors on a neuron. **a**, Surface representation of H_{CC} domain of BoNT/B (light blue) in a complex with Syt-II recognition domain (magenta) with a modelled G_{T1b} heptasaccharide (spheres) docked to the BoNT/B ganglioside binding site, modelled by proteins deposited in the PDB under accession numbers 1F31 (ref. 9) and 1FV3 (ref. 21). A cartoon model is shown to indicate the orientation of the heptasaccharide G_{T1b}. **b**, Model of BoNT/B binding using both Syt-II and ganglioside receptors at presynaptic membrane. The cytoplasmic domain of synaptotagmin is presented as two C2 domains (light pink) bound with calcium ions (yellow spheres), modelled using the coordinates of proteins 1TJX²⁸ and 1DQV²⁹ in the PDB.

Co-crystallization experiments using sialyllactose, a trisaccharide analogue of G_{M1} , as well as mutagenesis studies, suggest that there is a single ganglioside binding site within H_{CC} of BoNT/B, which is conserved among all clostridial neurotoxins^{9,17,18,20}. The ganglioside binding site is about 15 Å away from the synaptotagmin binding groove of BoNT/B. The sialic acid sugar (Sia6) of sialyllactose is buried in a pocket formed by residues of BoNT/B that share high conservation with the other serotypes of clostridial neurotoxins (Supplementary Fig. 5c). The equivalent site in H_{CC} of tetanus neurotoxin was shown to bind a G_{T1b} analogue containing all seven sugars, with the ceramide acyl chain replaced by a trimethylsilyl moiety²¹. The observation of ganglioside (G_{T1b} or G_{D1a})-facilitated BoNT/B binding to synaptotagmin hints at the potential function of the terminal Sia5 group and cooperative assembly of a BoNT/B–synaptotagmin– G_{T1b} ternary complex.

In an attempt to assess how BoNT/B recognizes both ganglioside and Syt-II, we docked the G_{T1b} analogue into the ganglioside binding pocket of BoNT/B by overlaying the corresponding Sia6 group of G_{T1b} (Supplementary Fig. 6) to that of sialyllactose in complex with BoNT/B (PDB accession number 1F31⁹). Interestingly, the longer branch of G_{T1b} extends the critical Sia5 group, the one that is only presented in G_{T1b} or G_{D1a} , towards I58 and N59 of Syt-II (Fig. 4a). This suggests that the previously identified ganglioside binding pocket in the BoNT/B–sialyllactose structure complex may actually represent only a portion of the recognition site, with the actual ganglioside binding sites being more extended. Indeed multivalent binding with larger oligosaccharides is a characteristic feature found in trefoil–carbohydrate recognition of bacteria to achieve higher avidity and selectivity²². If, as we anticipate, receptor binding sites of BoNT/B are linked energetically, the potency of a small-molecule inhibitor that targets a cooperative binding region may be amplified by blocking both Syt-II and ganglioside binding. Nonetheless, given the inherently flexible and less predictable nature of the oligosaccharide structure, obtaining a ternary structure of the BoNT/B–Syt-II– G_{T1b} complex is required to address this model.

The structure of the BoNT/B–Syt-II complex with modelled ganglioside discloses an enlightening molecular snapshot of BoNT/B while anchored to the presynaptic membrane² (Fig. 4b). When both ganglioside and Syt-II binding are presented, the H_{CC} domain of BoNT/B is locked onto the cell surface at one end by the two anchor points. Meanwhile, the translocation domain (H_N) of BoNT/B is positioned tangential to the cell membrane; a conformation that may assist pore formation on the synaptic vesicle to release the light chain. In addition, loop 1250 is shown to be close to the cell surface. It is possible that the side chains of Y1243 and Y1252 on the loop insert into the membrane, also facilitating pore formation. Our study presents a structural basis for the ‘double receptor’ hypothesis, and will also help in the development of inhibitors that may block binding of toxins to cell surface receptors.

METHODS

See Supplementary Information for detailed methods including Syt-II construct generation, pull-down assays, PC12 binding and entry assays, and ganglioside depletion/addition assays.

Protein preparation and crystallization. BoNT/B was dialysed extensively before crystallization as described²³. BoNT/B (8 mg ml⁻¹ in 50 mM HEPES buffer, pH 7.0, 100 mM NaCl, and 10 mM DTT) was incubated with the Syt-II recognition domain peptide at a 1:10 molar ratio and crystallized via the vapour diffusion method at 297 K. Plate-shaped crystals grew in a crystallization solution that contained 0.4 M (NH₄)₂SO₄, 0.9 M Li₂SO₄, and 0.1 M sodium citrate at pH 5.6. Crystals were transferred to mother liquor solutions of increasing concentrations of Li₂SO₄ (up to 90%) before harvesting and flash-cooled in liquid nitrogen.

Data collection and structure determination. Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 1-5 at a wavelength of 1.000 Å and processed using the HKL2000 program²⁴. The structure was determined by molecular replacement using the BoNT/B structure (PDB accession code 1SOE²³) as the search model with the program

MOLREP²⁵. Model building of the BoNT/B–Syt-II complex was done with Coot²⁶ and refinement with REFMAC5²⁷ using six TLS (translation, libration, screw) groups. Data collection and refinement statistics are summarized in Supplementary Table 1. Electron density was not observed for residues 40–44 and 60 of Syt-II or residue 441 at the light chain/heavy chain cleavage site of BoNT/B.

Received 9 August; accepted 3 November 2006.

Published online 13 December 2006.

1. Simpson, L. L. Identification of the characteristics that underlie botulinum toxin potency: implications for designing novel drugs. *Biochimie* **82**, 943–953 (2000).
2. Montecucco, C. How do tetanus and botulinum toxins bind to neuronal membranes? *Trends Biochem. Sci.* **11**, 315–317 (1986).
3. Mahrhold, S., Rummel, A., Bigalke, H., Davletov, B. & Binz, T. The synaptic vesicle protein 2C mediates the uptake of botulinum neurotoxin A into phrenic nerves. *FEBS Lett.* **580**, 2011–2014 (2006).
4. Dong, M. et al. SV2 is the protein receptor for botulinum neurotoxin A. *Science* **312**, 592–596 (2006).
5. Nishiki, T. et al. The high-affinity binding of *Clostridium botulinum* type B neurotoxin to synaptotagmin II associated with gangliosides GT1b/GD1a. *FEBS Lett.* **378**, 253–257 (1996).
6. Dong, M. et al. Synaptotagmins I and II mediate entry of botulinum neurotoxin B into cells. *J. Cell Biol.* **162**, 1293–1303 (2003).
7. Rummel, A., Karnath, T., Henke, T., Bigalke, H. & Binz, T. Synaptotagmins I and II act as nerve cell receptors for botulinum neurotoxin G. *J. Biol. Chem.* **279**, 30865–30870 (2004).
8. Nishiki, T. et al. Binding of botulinum type B neurotoxin to Chinese hamster ovary cells transfected with rat synaptotagmin II cDNA. *Neurosci. Lett.* **208**, 105–108 (1996).
9. Swaminathan, S. & Eswaramoorthy, S. Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B. *Nature Struct. Biol.* **7**, 693–699 (2000).
10. Breidenbach, M. A. & Brunger, A. T. Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* **432**, 925–929 (2004).
11. Simpson, L. L. Identification of the major steps in botulinum toxin action. *Annu. Rev. Pharmacol. Toxicol.* **44**, 167–193 (2004).
12. Jayaraman, S., Eswaramoorthy, S., Kumaran, D. & Swaminathan, S. Common binding site for disialyllactose and tri-peptide in C-fragment of tetanus neurotoxin. *Proteins* **61**, 288–295 (2005).
13. Lacy, D. B., Tepp, W., Cohen, A. C., DasGupta, B. R. & Stevens, R. C. Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nature Struct. Biol.* **5**, 898–902 (1998).
14. Schiavo, G., Matteoli, M. & Montecucco, C. Neurotoxins affecting neuroexocytosis. *Physiol. Rev.* **80**, 717–766 (2000).
15. Kitamura, M., Takamiya, K., Aizawa, S. & Furukawa, K. Gangliosides are the binding substances in neural cells for tetanus and botulinum toxins in mice. *Biochim. Biophys. Acta* **1441**, 1–3 (1999).
16. Kozaki, S., Kamata, Y., Watarai, S., Nishiki, T. & Mochida, S. Ganglioside GT1b as a complementary receptor component for *Clostridium botulinum* neurotoxins. *Microbiol. Pathol.* **25**, 91–99 (1998).
17. Rummel, A., Mahrhold, S., Bigalke, H. & Binz, T. The H-CC-domain of botulinum neurotoxins A and B exhibits a singular ganglioside binding site displaying serotype specific carbohydrate interaction. *Mol. Microbiol.* **51**, 631–643 (2004).
18. Ginalska, K., Venclovas, C., Lesyng, B. & Fidelis, K. Structure-based sequence alignment for the beta-trefoil subdomain of the clostridial neurotoxin family provides residue level information about the putative ganglioside binding site. *FEBS Lett.* **482**, 119–124 (2000).
19. Yowler, B. C., Kensinger, R. D. & Schengrund, C. L. Botulinum neurotoxin A activity is dependent upon the presence of specific gangliosides in neuroblastoma cells expressing synaptotagmin I. *J. Biol. Chem.* **277**, 32815–32819 (2002).
20. Rummel, A., Bade, S., Alves, J., Bigalke, H. & Binz, T. Two carbohydrate binding sites in the H-cc-domain of tetanus neurotoxin are required for toxicity. *J. Mol. Biol.* **326**, 835–847 (2003).
21. Fotinou, C. et al. The crystal structure of tetanus toxin Hc fragment complexed with a synthetic GT1b analogue suggests cross-linking between ganglioside receptors and the toxin. *J. Biol. Chem.* **276**, 32274–32281 (2001).
22. Weis, W. I. Cell-surface carbohydrate recognition by animal and viral lectins. *Curr. Opin. Struct. Biol.* **7**, 624–630 (1997).
23. Eswaramoorthy, S., Kumaran, D., Keller, J. & Swaminathan, S. Role of metals in the biological activity of *Clostridium botulinum* neurotoxins. *Biochemistry* **43**, 2209–2216 (2004).
24. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1997).
25. Vagin, A. & Teplov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
26. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
27. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Cheng, Y. et al. Crystallographic identification of Ca²⁺ and Sr²⁺ coordination sites in synaptotagmin I C2B domain. *Protein Sci.* **13**, 2665–2672 (2004).

29. Sutton, R. B., Ernst, J. A. & Brunger, A. T. Crystal structure of the cytosolic C2A–C2B domains of synaptotagmin III. Implications for Ca^{2+} -independent SNARE complex interaction. *J. Cell Biol.* **147**, 589–598 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Abola, K. Saikatendu and J. Ng for discussions and A. Walker for assistance with manuscript preparation. This work was supported by a grant from the Pacific Southwest Regional Center of Excellence (R.C.S. and E.A.J.), and grants from the NIH/NIAID (to E.R.C.). E.R.C. and E.A.J. acknowledge membership of and support from the Region V “Great Lakes” Regional Center of Excellence in Biodefense and Emerging Infectious Diseases. E.R.C. is an Investigator of the Howard Hughes Medical Institute. Portions of this research were carried out at the SSRL, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy

Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

Author Contributions J.W.A., Q.C. (crystallography) and M.D. (molecular biology) contributed equally to this work and co-wrote the paper. All authors discussed the results and commented upon the manuscript.

Author Information Atomic coordinates and experimental structure factors of the BoNT/B–Syt II ectodomain complex have been deposited in the RCSB Protein Data Bank (PDB) under accession code 2NPO. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.C.S. (stevens@scripps.edu) or E.R.C. (chapman@physiology.wisc.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Last week's meeting of the American Society for Cell Biology in San Diego, California, heard some tough words from the director of the National Institutes of Health (NIH), Elias Zerhouni. Using a phrase he has repeated often in speeches this year, Zerhouni began his keynote talk with the words: "Never before have so many novices faced so many disincentives to entering or continuing a research career." It was a phrase that resonated with many in the audience. But, as Zerhouni revealed, these words were not his own — they were written in 1982 by William Raub, who was at that time associate director for research and training at the NIH.

Zerhouni nevertheless agreed that times remain tough for life scientists in the United States — despite the fact that the NIH's budget is at a record high of some \$28 billion, after doubling between the years 1998 and 2003. But since that massive leap, the agency, which funds more than half of all US life-sciences research, has seen its budget stay flat — or even dip below inflation — with much of the money being taken up by grants funded at the beginning of the doubling cycle.

Worse still, the bigger budget seems, in some ways, not to be going as far as it once did. The amount of money requested for many biomedical grants has increased dramatically; reflecting the greater size and complexity of such research. Meanwhile, the success rate for applications has actually declined, a phenomenon Zerhouni calls "post-NIH-doubling fatigue syndrome".

The situation is especially troubling for young scientists — those who have the most difficulty landing the awards that would help them achieve autonomy. Zerhouni began addressing the problem with a new scheme called 'Pathway to Independence' — the first grants in this were announced last month. The programme funds fellows for two years under a mentor and three years once they are on their own. If the NIH does indeed give 150–200 such awards a year over the next five years, as Zerhouni intends, there should be less of a disincentive for young scientists to pursue research. Pushing NIH funding above inflation would help, too.

Paul Smaglik, *Naturejobs* editor

Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

ARE YOU READY FOR AN EXCITING CAREER THAT COULD HELP IMPROVE MILLIONS OF LIVES AROUND THE WORLD?

Then consider joining the scientific and medical forces at the National Institute of Allergy and Infectious Diseases (NIAID). NIAID, the second largest institute of the world-renowned National Institutes of Health (NIH), supports and conducts basic, applied and clinical research to better understand, treat, and prevent infectious, immunologic, and allergic diseases. The Division of Microbiology and Infectious Diseases (DMID) is an extramural research Division of NIAID. The Division supports extramural research related to the control and prevention of diseases caused by virtually all human infectious agents (over 250 pathogens) except HIV. DMID is seeking applications from exceptional candidates for the following position:

Deputy Director, DMID

As an extramural research Division, DMID supports a wide variety of projects involving various human pathogens and ranging from basic to clinical research, through the development and evaluation of new drugs, vaccines, and diagnostics. DMID-supported research involves many scientific areas including bacteriology and mycology, biodefense, emerging infectious diseases, enteric diseases, microbial genomics, parasitology, respiratory diseases, sexually transmitted infections, vaccine development, and virology. The Division also participates in global partnerships and maintains a drug development program.

The selected candidate serves as Deputy to the DMID Director and fully shares responsibility for managing DMID's national / international extramural research program. In collaboration with the Director, the Deputy fully participates in both the planning, administration, development, and evaluation of DMID's research programs, and in directing, overseeing, and evaluating DMID's on-going activities through subordinate supervisors, projects leaders, program specialists, and contractor employees. The Deputy represents the Division and the Institute at national and international meetings and conferences; makes scientific/clinical presentations at scientific/medical meetings; and serves as a key scientific advisor to DMID Branch Chiefs, Office Directors and NIAID. The chosen candidate has the ability to encourage participation and partnering with public and private/commercial entities to transition basic research knowledge into the development of products that will improve human health in the U.S. and globally.

Applicants must possess an M.D., Ph.D., or equivalent degree, be a U.S. Citizen, and exhibit a broad scientific vision; an ability to lead a staff; and a demonstrated expertise in managing a broad and complex biomedical research program. The chosen candidate will possess the interpersonal and communication skills required to interact effectively with representatives of academia, private industry, national and international research and health organizations, the media, and the general public.

This vacancy is being advertised under the Title 5 and Title 42 hiring authorities. Salary is commensurate with experience and a full package of benefits is available including retirement, health and life insurance, long term care insurance, leave and savings plan (401K equivalent).

For the Title 5 vacancy, applicants must be a U.S. citizen. To apply for this vacancy, please visit <http://usajobs.opm.gov>. Specific application procedures apply. Vacancy Announcement Number: NIAID-07-162175-DE, Supervisory Research Program Manager, GS-601-15, Open: 12/5/05 – 1/26/07.

For the Title 42 vacancy, Non-citizens may apply. Please send your curriculum vitae and bibliography to: Pamela McInnes, DDS, MSc(Dent), Director Center for Integrative Biology and Infectious Diseases (CIBID), NIDCR/NIH, Building 45/4AN12B, Bethesda, MD 20892. Direct inquires to: Dr. Pamela McInnes via email: pmcinnes@nidcr.nih.gov or at 301-443-8618.

The deadline for receipt of all applications is **January 26, 2007**.

We invite you to explore our Institute and view other available opportunities at:

<http://healthresearch.niaid.nih.gov/ddn>



HELP US HELP MILLIONS



**National Institute of Allergy
and Infectious Diseases**

NW93082R

THE NIH IS DEDICATED TO BUILDING A DIVERSE COMMUNITY IN ITS TRAINING AND EMPLOYMENT PROGRAMS



OPPORTUNITIES @ NIH THE NATIONAL INSTITUTES OF HEALTH



THE NATIONAL INSTITUTES OF HEALTH

OPPORTUNITIES @ NIH

ARE YOU READY FOR AN EXCITING CAREER THAT COULD HELP IMPROVE MILLIONS OF LIVES AROUND THE WORLD?

Then consider joining the scientific and medical forces at NIAID. As part of the Division of Microbiology and Infectious Diseases (DMID) at NIAID, the Parasitology and International Programs Branch (PIPB) is responsible for planning and conducting programs of extramural research aimed at understanding the biology of protozoan and helminth parasites and their interaction with the human host as well as their vectors and intermediate hosts. DMID/PIPB has the following scientific opportunity available:

Medical Officer

As a Medical Officer, the selected candidate will provide leadership and scientific/medical expertise and guidance in the planning, development, implementation and evaluation of translational and clinical research concepts, projects and initiatives to appropriate advisory groups; identify opportunities and problem areas, research gaps and relevant program needs and make recommendations for and facilitate new research efforts, clinical studies, clinical trials or other initiatives; and communicate with grantees/con-

tractors, cooperative group members/representatives and others on policy interpretation, merit review and evaluation processes and procedures, and on decisions, concerns or other issues/matters of a medical/scientific nature. The selected candidate will also oversee and advise on translational and clinical studies of parasitic diseases.

In order to be considered for this position, applicants must possess an M.D. or D.O. and have experience in translational and/or clinical research to examine the causes, diagnosis, treatment and prevention of infectious diseases. Research experience in bacteriology, mycology, virology, parasitic and other tropical diseases, or vector biology is required. Experience with clinical studies including developing and/or writing protocols, managing clinical study sites, and evaluating adverse and serious adverse events is desirable. Investigational product development, Phase I through Phase III clinical trials experience and knowledge of regulatory controls is also desirable. Board certification or eligibility in Internal Medicine, Pediatrics, or Infectious Diseases is

highly desirable.

This vacancy is being advertised under the Title 5 hiring authority. Applicants must be a U.S. Citizen. Salary is commensurate with research experience and accomplishments. A full Civil Service package of benefits (including retirement, health, life and long-term care insurance, Thrift Savings Plan, etc.) is available.

To apply for this vacancy, please visit:

<http://usajobs.opm.gov>. Vacancy announcement number: NIAID-07-160850 DH - Medical Officer, GS-602-13/14 Salary: \$81,190 - \$118,828 plus a Physicians Comparability Allowance up to \$30,000 may be paid. Applications must be submitted to the Human Resource Specialist. For questions, you may call 301-402-0957 or e-mail apply4jobs@niaid.nih.gov. Deadline for receipt of applications is **March 9, 2007**.

We invite you to explore our institute and other opportunities at:

<http://healthresearch.niaid.nih.gov/dn>



HELP US HELP MILLIONS



National Institute of Allergy and Infectious Diseases



Tenure-track Physician Clinical Center/Nuclear Medicine Department

This position is located in The Warren G. Magnuson Clinical Center, Nuclear Medicine Department (NMD).

We are seeking a research-oriented physician for a possible tenure-track position. An M.D. or M.D./PhD with U.S. Nuclear Medicine Board certification and CT training is needed to provide diagnostic and therapeutic nuclear medicine procedures as well as to participate in clinical research protocols of the NIH Intramural Program. U.S. citizenship or permanent residency status is required.

Please submit your curriculum vitae, bibliography, and a letter describing your clinical, research, and management experience to: **Mrs. Veronica Olaaje, HR Specialist, DHHS, NIH, OD/CSD-E, 2115 E. Jefferson Street, Rm. 2B209 MSC-8503, Bethesda, MD 20892-8503. Phone: 301-435-4748. Email: volaje@mail.nih.gov.**

Salary is commensurate with experience. This appointment offers a full benefits package (including retirement, health, life and long term care insurance, Thrift Savings Plan participation, etc.). Application packages should be submitted as early as possible, but no later than **December 31, 2006**.

Selection for this position will be based solely on merit, without discrimination for non-merit reasons such as race, color, religion, sex, national origin, politics, marital status, sexual orientation, physical or mental handicap, age or membership or non-membership in an employee organization.



TENURE-TRACK INVESTIGATOR: The Department of Health and Human Services (DHHS), National Institutes of Health (NIH), announces a search for a tenure-track scientist, to lead the structural biology program at the Intramural Research Program (IRP), National Institute on Drug Abuse (NIDA). Candidates must have a terminal (MD and/or Ph.D.) degree. Candidates will be evaluated on (1) knowledge and experience in proteomics, lipidomics, glycomics, and their bioinformatics; (2) expertise in analytical chemistry, with a major emphasis on mass spectrometry and chromatography; (3) an understanding of mass spectrometric techniques applied to surface imaging, including MALDI, LDI, MALDI-Ion Mobility-oTOFMS - with UV and IR lasers and SIMS, particularly as applied to neuroscience and drug abuse; (4) knowledge of direct tissue imaging analysis of drug and biomolecular localization; (5) demonstrated ability to prepare biological tissue samples for quantitative and semi-quantitative spatial analysis of small and large molecules by laser and ion microprobe and laser capture microdissection; (6) application of these techniques, as well as ESI, Ion traps, Quadrupole and Q-TOF ESI, (GC, TLC and HPLC) /MS, to real-time monitoring of levels of molecules in blood, cerebrospinal fluid and saliva; (7) experience in correlating analytical results with three dimensional theoretical molecular modeling of physiological structure; and (8) successful collaboration with investigators from a wide range of disciplines. The successful candidate may be hired through the NIH Title 42(g) program at a competitive salary rate and with full Federal benefits. Candidates must submit a CV, a statement of proposed research objectives and goals of not more than three pages, and four (4) letters of recommendation (from noncollaborators) to: **Barry J. Hoffer, M.D., Ph.D., Scientific Director, NIDA, and Chief, Cellular Neurophysiology Section, Cellular Neurobiology Research Branch, IRP/NIDA, 5500 Nathan Shock Drive, Baltimore, Maryland 21224**. A copy of the terminal degree (with a certified English translation if foreign) should also be included. **Email: bhoffer@intra.nida.nih.gov**. Application materials must be received by COB on **January 31, 2007**. Applications received after that date will not receive consideration.

Hotdogs at the end of the world

So long, and thanks for all the sausages.

Jeff Crook

Adam considered the sky. It wasn't even blue any more, just a sort of washed-out grey. The tall, pale man opposite him thoughtfully chewed a bite of hotdog, the bitten remainder of which he held about mid-chest, just slightly above Adam's head. Adam noticed a spot of yellow mustard on the man's red tie. His suit was black, as were the sunglasses resting on top of his wavy blond hair. His almost-colourless blue eyes remained fixed in an unblinking stare that Adam still found somewhat unsettling.

"What you say is true," Adam continued, "but the conflict between theology and interstellar travel isn't just with the book of Genesis. There's Revelations to consider. Of course, Norse mythology also includes an end-times myth, Ragnarok and all that, but there are so few Odin-worshippers these days, it seems pointless to include them."

"Entirely," the Nordic-looking man mumbled around his hotdog.

"Uh... entirely what?" Adam asked.

"Pointless," the man said. Adam noticed that the spot of yellow mustard had disappeared from the man's tie.

"Wonderful technology," Adam muttered before continuing. "But as you know, UFO researchers always assumed that the government actively concealed the existence of extraterrestrials to protect the Church, referring of course to the question raised by Genesis — if there are extraterrestrial civilizations, why aren't they mentioned in Genesis? The Brookings Institute is supposed to have done the original sociological study that has guided disclosure policy since the middle of the last century."

"That's true," the Nordic man said.

"It is? Oh, yes, but in my opinion the answer to the Genesis question wouldn't bring down the Church, would it? One could always argue that Genesis never says aliens don't exist, either."

"Obviously," the Nordic man said. He finished his hotdog. Adam removed another one from the grill, bunned it, squirted a line of mustard down its pink and char-blackened length, and handed it over the grill to his guest.

"Obviously. So the real sticking point is Revelations," Adam said. "Do you want a beer or something?"

"Yes," the Nordic man said. Adam opened a longneck using the bottle-opener feature on his grill spatula. A shadow of a smile twitched the corners of the tall man's narrow mouth. "That's a useful device," he said as Adam handed him the beer.

"So, as I was saying, Revelations and all the other Armageddon philosophies," Adam continued. "I mean, if there are other worlds out there where a man... being... whatever... can be born, live and die without ever setting foot on Earth, that kind of pulls the end-timers' teeth, doesn't it? Without an end-times in which sinners are judged and the righteous rewarded, Western religion becomes rather pointless. God destroys the world — big deal. Sure, a few billion people die, but in the big picture, it's a minor occurrence. Planets explode every day, am I right? Whole star systems go nova, trillions of intelligent life forms wiped out in the twinkling of an eye, no matter how moral or immoral they are or were. It's physics, and a lack of sufficiently advanced technology to detect the impending Armageddon and/or to escape it by fleeing their doomed planet."

The Nordic man ate his hotdog and drank his beer without blinking. Adam noticed that the man's bottom lip (which was somewhat fuller than his almost-fleshless upper lip) tended to get sucked into the neck of the bottle when he drank, making a comical squeaking noise when he pulled it out.

Adam glanced at the cloudless grey sky, his voice rising as he tried not to laugh. "And since these end-times religions self-perpetuate by exerting control over their congregants by holding or withholding the metaphysical keys to heaven, take away the

end times and what are they left with?"

"Where do you purchase these?" the Nordic man asked, gesturing with his beer to the half-eaten hotdog in his hand.

"I make them myself. Secret family recipe," Adam said. "The beef is Kosher."

"They are delicious," the tall man said. "I'll mention them to the Minister of Culture."

Adam smiled. He felt his point had been made. The back door opened and a small, light-grey alien stepped outside. "Thanks, Adam," the alien said with obvious relief. "I was about to bust."

"Not a problem, Thraz. Did you remember to flush?" Adam asked.

"Oh. Sorry." The alien blushed a deeper shade of grey and glanced at the Nordic man, who was still examining the uneaten remainder of his hotdog.

"Don't worry. I'll get it,"

Adam said. "Your pee dissolved the toilet last time. Try explaining that to the plumber. You want another beer or something?"

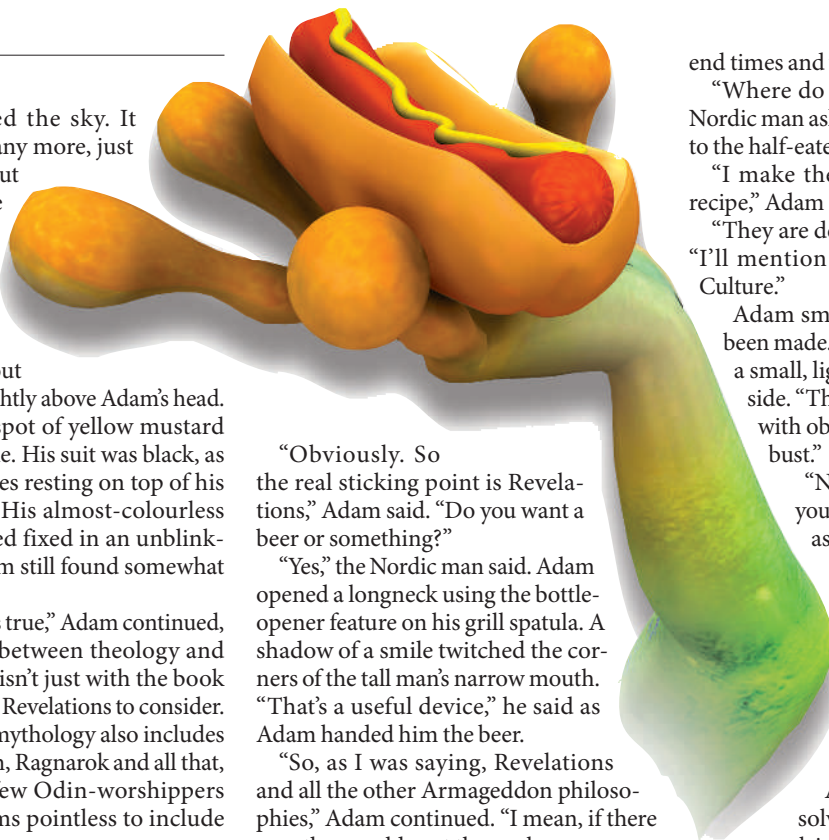
"Thanks, no. I'm driving. Well, see you later," the small alien said. He climbed into the metallic-green, lozenge-shaped spacecraft parked in Adam's backyard. The Nordic man ate the last half of his hotdog in one bite, set the empty beer bottle on a fencepost, and climbed into the spacecraft without so much as saying goodbye. Adam was used to this. The tall ones seemed unemotional, but the little guys were friendly enough.

Adam waved goodbye as the door materialized and blended seamlessly with the hull of the ship. The craft rose silently to treetop level, then shot off towards the encroaching sea, leaving behind a rainbow arc that shimmered in the washed-out sky for a second or two before fading. Adam smiled and popped an olive into his mouth as the sound of waves reached his ears.

High tide again, and getting closer every day. So much for that covenant.

Yes, interstellar travel is a lovely way to escape the end of the world and all its moral obligations, Adam thought with a rising sense of hope and purpose. ■

Jeff Crook, father of two, husband of one, author of many, publisher of few, editor, too, is halfway to the end of his squib: (n) — a noisemaker or failed magician.



JACEY